Vol. 51, No. 1 January, 1944

Psychological Review

EDITED BY
HERBERT S. LANGFELD
PRINCETON UNIVERSITY

CONTENTS

The Logic of Psychophysical Measurement: GUSTAV BERGMANN AND KENNETH W. SPENCE 1

Repetition and Learning—I. Stimulus and Response: THOMAS W. COOK 25

The Screen Test in Military Selection: WILLIAM A. HUNT, CECIL L. WITTSON, AND HERBERT I. HARRIS 37

The Nature of Theory Construction in Contemporary Psychology: KENNETH W. SPENCE 47

If-Then Relations in Paralogics: H. M. JOHNSON 69

Brightness Enhancement in Flickering Light: THEODORE LOUIS JAHN .. 76

PUBLISHED BI-MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

PUBLICATIONS OF

THE AMERICAN PSYCHOLOGICAL ASSOCIATION

WILLARD L. VALENTINE, *Business Manager*

PSYCHOLOGICAL REVIEW

HERBERT S. LANGFELD, *Editor*
Princeton University

Contains original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.
Subscription: $5.50 (Foreign, $5.75). Single copies, $1.00.

PSYCHOLOGICAL BULLETIN

JOHN E. ANDERSON, *Editor*
University of Minnesota

Contains critical reviews of books and articles, psychological news and notes, university notices, and announcements. Appears monthly (10 issues), the annual volume comprising about 665 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

S. W. FERNBERGER, *Editor, on Leave,* FRANCIS W. IRWIN, *Acting Editor*
University of Pennsylvania

Contains original contributions of an experimental character. Appears bi-monthly (beginning 1944), one volume per year, each volume of six numbers containing about 520 pages.
Subscription: $7.00 (Foreign, $7.25). Single copies, $1.25.

PSYCHOLOGICAL ABSTRACTS

WALTER S. HUNTER, *Editor*
Brown University

Appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

PSYCHOLOGICAL MONOGRAPHS

JOHN F. DASHIELL, *Editor*
University of North Carolina

Consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.
Subscription: $6.00 per volume (Foreign, $6.30).

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

GORDON W. ALLPORT, *Editor*
Harvard University

Appears quarterly, January, April, July, October, the four numbers comprising a volume of 550 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.
Subscription: $5.00 (Foreign, $5.25). Single copies, $1.50.

JOURNAL OF APPLIED PSYCHOLOGY

DONALD G. PATERSON, *Editor*
University of Minnesota

Covers the applications of psychology in business, industry, education, etc. Appears bi-monthly, February, April, June, August, October, and December.
Subscription: $6.00. Single copies, $1.25.

Subscriptions, orders, and business communications should be sent to

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

VOL. 51, NO. 1 JANUARY, 1944

THE PSYCHOLOGICAL REVIEW

THE LOGIC OF PSYCHOPHYSICAL MEASUREMENT

BY GUSTAV BERGMANN AND KENNETH W. SPENCE

State University of Iowa

I. INTRODUCTION

The aim of this paper is to present a methodological analysis of some of the problems which arise in connection with psychophysical measurement and, incidentally, of certain allied aspects of psychology such as perception and psychological measurement in general. The discussion falls into three main parts: In the first section an attempt is made to indicate the manner in which scientific empiricism, the standpoint from which the present survey was undertaken, approaches these problems. Our exposition of this epistemological point of view must necessarily be restricted to a few very cursory remarks which have an immediate bearing on the problems at hand.[1]

The second section, which deals with physical measurement, cannot pass as a systematic exposition either. Its main purpose, again, has been to call attention to those points concerning physical measurement which must be understood in order to secure an adequate grasp of psychophysical measurement and of psychological measurement in general. Readers who find this second section difficult are referred to the elementary chapters on physical measurement which can be found in several introductory textbooks on logic and scientific method (**7, 8, 9, 13**). An interesting paper by

[1] For a more complete and systematic account of scientific empiricism (logical positivism) the reader is referred to the writings of Carnap, Morris, and others, particularly to the monographs collected in the *International Encyclopedia of Unified Science* (**21**).

Nagel (12) is on the intermediate level while the most elaborate analysis of physical measurement so far is due to Campbell (6).[2]

In the third and last part of the paper the particular problems of psychophysical measurement have been reexamined in the light of the points previously elaborated. Also included as a necessary background for this discussion is a sketch of the status of perception as conceived by the behavioristic psychologist.

II. The Methodological Frame of Reference

1. Scientific empiricism does not hold to the view, rightly or wrongly associated with Watson's name, that so-called mentalistic or introspectionistic terms such as 'sensation,' 'consciousness,' 'image,' etc., are *necessarily* meaningless or that their referents do not exist. But scientific empiricism does hold that it is the methodological ideal of the sciences of behavior to use such mentalistic terms only after they have been introduced by (*operational*) *definitions from a physicalistic meaning basis.* The meaning of the italicized portion of this statement can probably best be made clear by first giving illustrations of concepts or terms employed in science which meet this criterion. Following this we shall attempt to elucidate further by means of certain simple instances. Familiar examples of satisfactorily defined concepts are 'weight,' 'mass,' 'electric conductivity,' etc., in physics, and 'I.Q.,' 'stimulus trace' (Hull) and 'habit strength' (Hull) in psychology. The following paragraph may, for our immediate purpose, pass as an elucidation.

When mentalistic terms are introduced by (operational) definitions from a physicalistic meaning basis, then every statement which contains such terms can be tested by the *scientist's* observations (unquantified) and measurements (quantified) of physical objects. The following are three examples of sentences which express such basic observations and measurements of the scientist: 'object *A* is green,' 'the

[2] The extent to which the methodological analysis of physical measurement has influenced psychology is best revealed in two papers, an earlier one by McGregor (11) and a recent analysis of mental tests as measuring instruments by Thomas (19).

pointer of instrument *B* coincides with the point *a* of its scale,' 'the density of object *C* is 3.57.' The first of these statements is, in a certain sense, simpler than the second, because the confirmation of the latter involves the recognition of the instrument *B* and thus the confirmation of several statements like the first; again, the confirmation of the third presupposes the confirmation of such statements as the first and the second. The theory of physical measurement or, what amounts almost to the same thing, the operational analysis of physics consists in the reduction in this manner of even the most abstract of its concepts or, as one also says, empirical constructs. Reduction of a concept thus involves indicating the classes of statements as "simple" as our first example, the verification of which is considered a confirmation of the statements in which the concept occurs.

2. Scientific empiricism holds to the position that all sciences, including psychology, deal with the same events, namely, the experiences or perceptions of the scientist himself. The behavior scientist who claims to study such perceptual behavior in his subjects is thus asked to start uncritically from his own perceptions; a piece of advice, the apparent circularity of which has continued to puzzle many of the more philosophically-minded psychologists. To them neither physicalism, a less sophisticated form of empiricism, which dogmatically outlaws all attempts to clarify the puzzle of this circularity by branding them as revivals of the old mind-body metaphysics, nor solipsism the other, more traditional council of despair is acceptable. Consequently the philosophical uneasiness thus created appears to find expression in a variety of quasi-metaphysical speculations.

In the schema outlined by the scientific empiricist the experiences of the observing scientist do indeed have a privileged, even unique position. If pressed too far and without the necessary epistemological sophistication, this account of the scientist's position can very easily lead to a metaphysical thesis of the solipsistic type. The fact of the matter is that science is always epistemologically dogmatic. Scientists, *qua* scientists, are usually not interested in such

matters beyond receiving the philosopher's assurance that their methodological schema does not allow for any undesirable epistemological inferences. Nevertheless, attention may be called here to the re-examination of the mind-body issue offered by scientific empiricism (2). This reëxamination tries, on the one hand, to overcome the dogmatism of *epistemological* physicalism, and on the other, to justify the *methodological* solipsism and physicalism of science. So far as the psychologist is concerned these remarks may be expressed in a somewhat different manner. They are to the effect that the empiricist scientist should realize that his behavior, symbolic or otherwise, does not lie on the same methodological level as the responses of his subjects, and consequently that he should not in reporting the latter use any mentalistic terms which have not been introduced from a physicalistic meaning basis. This latter fact may also be expressed by saying that all his terms should be behavioristically defined.

3. In the terminology now prevailing among empiricists the last point can be stated in the following way: In studying his subjects, including their symbolic responses (object language), the behavior scientist himself uses a different language (pragmatic metalanguage). That linguistic behavior can be utilized in operational definitions, even according to the most rigorous standards, is most generally recognized today. It is important in psychology, however, to distinguish between such utilization of symbolic responses of human *subjects* on the one hand and the use of responses of human *judges* on the other. Let us consider a schematically simplified example. Assume that a scientist calls a behavior fragment of a subject aggressive if 8 out of 10 judges apply this word to the behavior in question. In doing so our scientist has made use of what might be termed the *human yardstick* in the introduction of *his* term 'aggressive.' The latter lies in the scientist's metalanguage and must, therefore, be carefully distinguished from the judges' term which belongs to the object language. The remarkable statistical refinement which has recently been brought to bear on definitions

of this kind has tended to overshadow the rather fundamental methodological difference between concepts introduced with and those introduced without the use of the human yardstick. If the latter are called *strict behavioristic terms*, one would have to say that the scientist of our illustration describes, in strict behavioristic terms, not the behavior of his subjects, but rather the behavior of the group composed of his subjects *and* of his judges.

Since 'behavioristic' has become an honorific label, it might be useful to add that the preceding remarks are not meant as a methodological criticism of correctly defined, but not strict behavioristic terms; they are merely a clarification of their peculiar status. The indispensability of such terms in present day clinical and social psychology is readily conceded. The clinician who applies mentalistic, not strictly introduced, terms to his patients is indeed the limiting case of the schematic situation; in this instance the single judge coincides with the observing scientist. There is, of course, no objection against calling all these procedures behavioristic, but then the word refers to nothing more precise than what is usually called the experimental attitude. And it is only fair to say that the theoretical psychologist as well as the empiricist philosopher are nowadays mainly interested in a behavior science which is behavioristic in the stricter sense (see below, IV 4 and V).

The bearing of these considerations upon some of the quantitative methods most widely used in psychology will become clear in the course of the discussion. And they raise also a question concerning psychophysical measurement proper. Does a human observer who discriminates, in one of the traditional situations, between, say, two pitches, serve as a judge or as a subject? The answer, it will be seen, depends upon which aspects of the psychophysical situation are emphasized.

III. Physical Measurement

1. Some properties of and relations between physical objects or events are accessible to further differentiation.

The meaning of this statement can only be illustrated. Loudness, hotness, length can be differentiated in this sense; the property 'male' ('X is a male') and the relation 'to-the-left of' ('A is to the left of B') cannot. Properties of physical objects or events which are amenable to such differentiation we shall call physical *dimensions*. Physical measurement consists in the assignment of numbers to the objects or events of a physical dimension in accordance with certain rules (laws and conventions). The result is a *scale*. The distinction between a physical dimension and any of the scales which have been developed to measure it, pedantic as it might seem, is necessary in order to avoid certain confusions which have arisen. For instance in discussions on measurement a distinction has sometimes been made between intensive and extensive scales. It will be shown that intensity and extensity are properties of dimensions and should not be used with reference to scales.

2. A dimension is said to be *intensive* if the following three conditions are fulfilled. *First*, that there exist two physical relations, designated by '$\geq$' and '$=$' respectively, so that the statements '$X \geq Y$' and '$X = Y$' are both meaningful, X and Y being any two objects or events which have the property that defines the dimension. *Second*, that these statements can be tested by the scientist's observations or *manipulations within the dimension*. The italicized clause will be explained presently (III 4). *Third*, that the two relations fulfill certain axioms, the so-called axioms of order (**12**, p. 315, 1–6). These axioms are empirical laws, *i.e.*, whether they hold or not for any two physical relations is entirely a question of fact. Two of these axioms, the intuitive meaning of which is obvious, are given here for the purpose of illustration.

(1) Either $X = Y$, or $X \geq Y$, or $Y \geq X$.

(2) If $X \geq Y$, and $Y \geq Z$, then $X \geq Z$,

where X, Y, Z, are arbitrary objects or events of the dimension.

If these three conditions are fulfilled, then it is (in more than one way) possible to assign to each object or event X

of the dimension a number designated by $N(X)$, so that

$$N(X) > N(Y) \quad \text{if and only if} \quad X \geq Y,$$
$$N(X) = N(Y) \quad \text{if and only if} \quad X = Y.$$

Here '$>$' and '$=$' are the usual arithmetical symbols. If such a coördination has been made then a rank order scale has been constructed for the dimension.

A dimension is called *extensive* if two additional conditions are fulfilled. *Fourth*, that there is a further physical operation within the dimension, designated by '$+$,' which is of such a nature that if performed upon any two objects or events of the dimension, it results again in an object or event of the dimension. This latter object or event is designated by '$X + Y$.' *Fifth*, this operation, too, is supposed to fulfill certain axioms (**12**, p. 315, 7–12).[3] Two of them, here given for the purpose of illustration, are

(3) If $X = X'$ and $Y = Y'$, then $X + Y = X' + Y'$.

(4) $X + Y = Y + X$.

Furthermore, in an extensive dimension it is always possible to assign the numbers to the objects in such a way that

(A) $N(X + Y) = N(X) + N(Y)$.

Again '$=$' and '$+$' are the usual arithmetical symbols and are not to be confused with '$=$,' '$\geq$,' '$+$,' the names of physical relations and operations. A scale which has this property (A) is called *additive*. That is to say, extensive dimensions and, by the very nature of this definition, only extensive dimensions allow for the construction of additive scales.

Perhaps a simple illustration will show the value of this cumbersome terminology. Length, as is well known, is extensive and its usual scale is additive. If, however, one were to replace the numbers of our ordinary yardsticks by those of which they are the logarithms, the axioms of order

[3] It is not the purpose of this paper to examine the formal adequacy, independence, etc., of the set of axioms which Nagel takes over from the mathematician Hoelder.

and extensity would still be fulfilled, but instead of equation (A) one now has $N(X \underline{+} Y) = N(X).N(Y)$. The ordinary slide rule is another such non-additive length scale. This is one of the reasons why we distinguish between dimensions and scales, between extensity and additivity.

3. One of Campbell's and Nagel's examples (12) will be used to illustrate the next point.[4] The example deals with the density of liquids. This construct is so defined that liquid Y is said to be of greater density than liquid X if X floats on top of Y. (One of our simplifying assumptions is that the two liquids do not mix.) With this definition, density turns out to be an intensive dimension. The physical relation, '$X \underline{\geq} Y$,' reads in this case 'Y floats on top of X' and it will be observed that an operation '$\underline{+}$' cannot be found. Let us now attribute numbers, designated by $D(X)$, to this intensive dimension. As we know, this can be done in a great many different ways; the only requirement is that the greater ($\underline{\geq}$!) density gets the greater ($>$!) number. In actual practice, of course, density is not defined in this way, but as the quotient of weight (W) and volume (V). Let us designate this new 'density' of a body X by a lower case symbol: $d(X) = W(X)/V(X)$. If one says that $d(X)$ and $D(X)$ are two different measures of the "same" construct one has reference to the following empirical laws:

$$d(X) > d(Y) \quad \text{if and only if} \quad D(X) > D(Y),$$
$$d(X) = d(Y) \quad \text{if and only if} \quad D(X) = D(Y).$$

Because of these empirical laws which obtain between the three dimensions, volume, weight, and density (D), one can use $d(X)$ instead of $D(X)$ as a scale of 'density.' If, in order to scale a dimension, one thus makes use of other dimensions and of the empirical laws obtaining between them and the dimension one wants to scale, then the resulting scale is called a *derived* scale. The usual temperature scale is another familiar instance of a derived scale for an intensive dimension, the extraneous dimension involved being volume, the em-

[4] It must be borne in mind that all our illustrations are extremely simplified and not descriptions of practical working procedures.

pirical law used that of thermic expansion. By means of derived scales one can also quantify or "measure" dimensions which are not even intensive in the sense defined above, that is to say, where no physical relations *within the dimension* that fulfill the order axioms can be found.

4. We are now in a better position to elucidate the clause 'manipulations (relations) within a dimension' to which attention was called above (III 2). Manipulations within a dimension do not involve utilization of any of the empirical laws which connect it with other dimensions. The manipulations, for instance, by which we compare the lengths of yardsticks and build their operational sum, '$\underline{+}$,' lie entirely within the dimension of length. That is to say, the empirical laws which make length an extensive dimension are the axioms of extensity, if the symbols '$\underline{=}$,' '$\underline{>}$,' '$\underline{+}$' are read with the appropriate interpretations; *e.g.*, '$X \underline{+} Y$' reads "Take the stick Y and put one of its ends as close as possible to one end of the stick X, etc." And these laws have no reference to any other dimension.

It should be clear from the above discussion that only in the case of fundamental scales, *i.e.*, those based on operations within the dimension, is inference as to the intensity or extensity of the dimension possible. Derived scales do not allow any such inference. To say the same thing still differently: A dimension is shown to be intensive or extensive respectively, if two or three operations, '$\underline{=}$,' '$\underline{>}$,' '$\underline{+}$,' within the dimension have been found and shown to fulfill the 6 or 12 axioms.

5. Everybody agrees that the customary density scale is preferable to an arbitrary rank order; that it is more convenient to measure time in seconds than by counting the heart beats of a totem animal, to quote an illustration which Schlick was fond of using in his lectures. What, however, is the correct analysis of this situation? Is it because the one scale measures the *true* density; because any two seconds are *equal* while the intervals of a heart beat are not? Or, to mention still another expression sometimes ambiguously used in this context, is it the case that only one of the scales from

which we can choose is *valid?* To the confusion surrounding these questions our attention must now be turned.

Consider the law of thermic expansion. Using the customary temperature scale this law can be formulated in the following manner: the volume of any substance is (within certain limits, approximately) a linear function of its temperature: $v_t = v_0(1 + c_X t)$. Let us call 'L' this formulation of the law. Assume now that temperature has been scaled by an arbitrary rank order (R), so that, as we know but as the users of the scale do not know, $1R = 0°$ C., $2R = 1°$ C., $3R = 4°$ C., $4R = 17°$ C., etc. Designate, furthermore, by $c(X)_{1,2}$, $c(X)_{1,3}$, $c(X)_{1,4}$, etc., the percentual increase in volume (expansion coefficient) of the substance X between the temperature levels indicated by the indices. Experiments would then lead to the formulation of the following law (L_1): the expansion coefficients of any substance stand in the following relations: $c_{1,3} = 4c_{1,2}$, $c_{1,4} = 17c_{1,2}$, etc.

L_1 is of course the form which L would take if one works with our unusual rank order. One sees that L_1 is mathematically more complicated and less intuitive than L. This observation leads to the very core of the matter under discussion. We choose our scales so that certain empirical laws receive an expression as intuitive and (or) as mathematically simple as possible. It is an inaccurate and misleading way of speaking when such choice of scales is described as an attempt to equalize the unit distances of the scale. If there is a choice as to which law should be selected, then one usually picks that which one considers more "fundamental." But one must realize that 'fundamental' is again but a figure of speech; those laws are fundamental which play or which we expect to play a prominent rôle in our everchanging theoretical structure. If this structure changes or expands, then we frequently also change our scales. In the case of temperature, for instance, the mercury scale has been replaced by the gas scale and the latter, in turn, at least for theoretical purposes, by the entropy scale. Each of these steps marked a progress in the sense that the temperature scale became more directly linked to more and more fundamental laws

within an expanding theoretical structure. But it is merely another figure of speech to describe these successive steps as approximations to a true or valid scale of temperature. In itself no scale is more valid than any other. It is hunting the will-o-the-wisp to believe that such refinement as just described will ever make a scale additive. Additivity, as we have seen, can not even be defined for nonextensive dimensions.

6. What, in general, is meant by the equality of any two units of a scale or, as one also says, by the equality of differences? An answer to this question has already been partially given in the preceding paragraphs; all that remains is to make it explicit and to supplement it by some concluding remarks. If a derived scale is built by utilizing an empirical law and if the assignment of numbers is made in a manner to give to this or another empirical law a mathematically simple expression, then the *numerical* identity of the differences between two pairs of scale points has *empirical* meaning. Sometimes, as in the case of the mercury scale for temperature or the usual time scale, this meaning is more or less intuitive, sometimes, as in the case of the entropy scale, the simplicity lies on the theoretical level. Always, however, the numerical equality of the differences between two pairs of scale points has exactly that much empirical meaning as has been put into the scale by taking account of empirical laws in attributing the numbers. There is no end to the possible refinements nor to the many quite different ways in which meaning can thus be given to a scale. But again it is very misleading if the search for such refinement, which is the search for empirical laws that could be formulated in any scale, is mistaken for the search for true or valid measurement; or, as happens to be the case in psychophysics, for the search for extensive or additive scales.

In the case of derived scales the empirical laws drawn upon are always of the form '$y = f(x)$,' where 'y' stands for the derived dimension to be scaled and 'x' for one or several extraneous dimensions already scaled. There are, however, other ways to endow a scale and, in particular, the equality of the numerical differences with empirical meaning. This

time we shall illustrate by means of the usual hardness scale which, as is well known, is an arbitrary attribution of the numbers 1 to 10 in an intensive dimension. Accordingly, there is no factual meaning in the statement that the hardness differences of two pairs of minerals with the numbers 8, 7; 3, 2, respectively, are equal. But let us now assume that we have ranked, with respect to hardness, all known minerals. Assume, furthermore, that there are altogether 901 of them, and that we rescale hardness so that the softest one gets the number 1, the 101 the number 2, and 201 the number 3, and so on. If such a scale is being used, is our statement about the four minerals (8, 7; 3, 2) still devoid of all factual meaning? Obviously not, for now it means that there are exactly as many minerals the hardness of which lies between those of the first pair as there are minerals the hardness of which lies between the second pair. There is admittedly not much point to this refinement, but the example has at least the virtue of simplicity.

A very common psychological scaling technique can be shown to rest upon exactly the same principles. Assume that a property is, in a statistically stable manner, distributed over a large population and that a percentile scale has been based, in the familiar fashion, upon the area under that curve. The question then arises as to what meaning can be attributed to the numerical equality of score differences, such as, let us say, between 60 and 50, and 25 and 15 respectively. Despite the opinion current among psychologists that this numerical equality does not signify 'equality,' it does have a well defined empirical meaning, namely this: in any fair sample the number of individuals between 60 and 50 is equal to the number of individuals between 25 and 15. The situation here is, in principle, the same as in our example of hardness. In every such instance some empirical law is used to endow the numerical equality of the differences between scale points with some kind of empirical meaning. In the case of the mercury scale for temperature the law drawn upon is the expansion formula, in the case of the percentile scale

it is the (assumed) stability of some statistical distribution, normal or otherwise.

In this connection it might be worthwhile to call attention to the correct analysis of what is often thought of as the equality of sigma units (*Z*-scores, *T*-scores, etc.). That a certain concept is normally distributed is a significant empirical law about that concept and for theoretical reasons psychologists are often interested in defining concepts which are thus distributed. However, the translation of raw scores into sigma units amounts, technicalities apart, to nothing but the multiplication with a constant factor. There is no methodological reason for the privileged status which many contemporary psychologists attribute to such units. On this point our analysis is in essential agreement with that of Thomas (19).

Let us close this section by a remark with which the discussion of this point often starts. In extensional dimensions and only *in extensional dimensions, a factual meaning can be given to the equality of the numerical differences without drawing upon any other empirical laws than the axioms of measurement themselves.* The fact is that wherever there is an operation '$\underline{+}$,' there is, in all known instances, also an inverse operation '$\underline{-}$,' so that the two operations stand, loosely speaking, in the same relation as the two arithmetical symbols '+' and '−.' And *if an additive scale is used,* then it is also true that

$$N(X) - N(Y) = N(X') - N(Y') \quad \text{if and only if} \quad X \underline{-} Y \underline{=} X' \underline{-} Y'.$$

The "empirical equation" to the right expresses the empirical meaning of the numerical equality to the left. It is a corollary of the preceding analysis that the phrase 'the units of a scale are equal by definition' is, to say the least, not very enlightening.

IV. Psychophysical Measurement

1. Psychophysical measurement has traditionally dealt with the quantification of *attributes* (psychological dimensions).

The neutral term 'quantification' instead of 'measurement' has been chosen purposely. *An attribute is not a property of physical events or objects, but is defined by means of the discriminatory responses to such objects or events on the part of observers who are different from the scientist.* Again, the term 'observer' has been chosen because of its neutrality; at the present stage of the discussion it should not be identified with either the 'subject' or the 'judge' of an earlier section (II 3). The customary definitions of attributes are all based upon the verbal discriminations of human observers or, what amounts to the same, upon their motor responses in accordance with verbal instructions. Attributes such as pitch and loudness are thus actually defined by the consistant use of one and the same adjective and its comparative forms. It is important to notice that such responses, even if they are verbal, do not belong to the (pragmatic meta-) language of the scientist, but to the (object) language of the observers. By means of their utterances such as 'equally loud' or 'louder,' the scientist introduces his empirical construct 'loudness.' To repeat the same thing in a slightly different form, one can also say that the operational basis for the introduction of an attribute includes that of the pertinent physical dimensions and, *in addition*, the observers and the procedures of eliciting their responses.[8] The situation is thus analogous to the one discussed with reference to the social psychologist's term 'aggressive,' and the observers take, in a certain sense, the place of the judges. But there is also an important difference which must not be overlooked. The observers respond to a physical dimension already scaled by physical measurement, while judges are employed where no working physical scale has yet been found.

These distinctions become particularly important if *numbers* or, rather, to adopt a convenient usage, *numerals* occur in the observer's responses, as in the method of fractionation where the observers are required to bisect or equipartition supraliminal intervals, or to adjust the variable stimulus to

[8] Attention should be called to the fact that by means of the conditioning technique attributes for nonarticulate observers can be defined.

'half the given pitch' (17, 18). In using such numerical estimates of observers for the construction of scales one is *not* using numbers in the same sense and in the same way as one does in physical measurement.[6] One only creates confusion if he uses terms like 'additivity' or 'multiplicativity' in referring to such response scales. Furthermore, confusion is enhanced by the failure to distinguish between additivity of scales and extensionality of dimensions. There shall be occasion to return to this point (IV 5). The distinction between the use of numerical estimates by judges or observers, on the one hand, and the scientist's direct, strictly behavioristic use of numbers, on the other, applies of course also to scalings by means of equal appearing intervals outside the field of the so-called sensory discriminations.

2. If one were asked to condense into one sentence a characterization of most of the *nonphysiological* work done in psychophysics, he might venture the following formulation: By laying down the operational procedures (method of limits, constant stimuli, fractionation, etc.), psychophysics first defines the various sensory attributes and then studies the empirical laws that obtain between the responses on the one hand and the dimensions of the physical objects or events, to which these responses are made, on the other. For instance, the procedures of any of the existing methods which lead to the establishment of a j.n.d.-scale contain the (operational) definition of the attribute in question. The *Weber-Fechner curve* thus obtained is one of the empirical laws of psychophysics. And there is of course no *a priori* reason why, for instance, the curve of "constant-stimuli loudness" should coincide with that of "fractionation loudness."

The adoption of the convenient term 'Weber-Fechner curve' requires two clarifications. *First*, by this curve is

[6] As the terms are used in this paper, numbers belong to the scientist's metalanguage, numerals to the object language used by the subject. This clarification also leads to the correct interpretation of the so-called *ratio scales*, mentioned by Boring (5, p. 52). It might at least be mentioned that an illegitimate fusion between these two ways in which 'numbers' occur plays an important rôle in Ramsey's unsuccessful attempt to construct a theory of probability (1). See also the last paragraph of IV 5.

meant the *empirical curve* representing the relationship between the quantification of an attribute, based on j.n.d.'s or any other procedure such as fractionation, on the one hand, and the pertinent physical dimensions on the other. Thus neither the logarithmic nor any other form of this empirical relationship is assumed. It would indeed be strange to make any *a priori* assumptions of this kind. *Second*, and in view of Stevens' results (15), the expression 'curve' should be replaced by 'polydimensional surface,' for as this investigator has shown in some instances, one and the same attribute is, as a matter of fact, a function of several physical dimensions. Pitch, for instance, has been discovered to be a function of both frequency and energy. The classical view of a one-to-one correspondence between physical and psychological dimensions thus turned out to be an oversimplification. The theoretical significance of this finding needs no emphasis. But since this refinement does not affect the present argument and since dealing with surfaces would make our formulations unnecessarily cumbersome, we shall go on speaking about the Weber-Fechner curve.

3. Assume now that by one of the psychophysical methods a Weber-Fechner curve between, let us say, pitch and frequency has been established. If this curve is monotonous, as it actually is, then one can use what we shall refer to hereafter as the *psychological* scale of j.n.d.'s or mels (17) as a scale of the physical dimension frequency. Such a procedure, though not always fruitful, is entirely legitimate and its very possibility throws an interesting light on the nature of psychological scales. As has already been pointed out, such "psychological" measurement can also be considered as a rescaling of physical dimensions by means of observer responses. The Weber-Fechner curve is the mediating empirical law.

It is realized that what is here called a psychological scale is referred to as a psychophysical scale within the traditional field of psychophysics and as a psychological scale only in such instances as the scaling of attitudes, in which no

independent physical measurement exists. So far as the instructions to the participating organism and its behavior are concerned, however, there is really no difference between the two cases. Furthermore, both have in common the fact that the experimenter is not interested in the response aspects themselves but rather in the stimulus dimension involved. The difference lies entirely in the circumstance that only in the traditional psychophysical case is there an independent physical scale of the stimulus dimension. In our discussion this difference is expressed by the distinction we make between the two terms 'judge' and 'observer.'

It is also worthwhile noticing in this connection that the empirical content of the Weber-Fechner law consists in the functional connection between the physical dimension and certain response aspects within the chosen procedures such as the number of j.n.d.'s.[7] If one recalls that derived scales are established by means of mediating empirical laws between *two* independently defined *physical* dimensions, then it becomes clear that the scales of traditional psychophysics do not fall into this category. If interest lies in the stimulus aspect then they represent an additional (psychological) scaling of a (physical) dimension already scaled. When the response aspect is emphasized, then we refer to them as scaling an attribute or psychological dimension.

Within the general field of measurement the scales of psychophysics thus constitute a type of their own. Continuing our analysis with a view to describing further operational differences between these scales and physical measurement it is appropriate to elaborate once again the inapplicability to them of the two categories 'intensive' and 'extensive.' Like 'derived' these two terms apply only to physical dimensions and not to attributes.

Where indeed are the *scientist's operations upon observer responses* which would justify their use? We must confess to utter inability to think of any operation '$\pm$,'

[7] Accordingly, the only correct derivation of the Weber-Fechner formula is that which does not make use of differentials. Such a derivation is to be found in Woodworth's textbook (30, p. 435–437).

which could be applied to the observers or to any of their responses which have so far been used for the introduction of attributes. Consider the case of loudness and assume even, contrary to fact (16), that the observers' judgments were 'twice as loud' if two equal sound emitters, instead of one, are put into action. In this case '$\underline{+}$' obviously means the simultaneous operation of the two sound emitting devices. What is thus operated upon is the physical dimension, not the attribute or the observer; and this in spite of the occurrence of numerals (ratios) in the observer reports. The same criticism holds true for the relation (operation) '$\underline{>}$' or intensity, even if one were to use the term 'operation' in a sense broad enough to include simple sensory discriminations. The fact would still remain that what falls, in a vague sense, into a serial order are the operations (discriminations) of the observer, not those of the scientist, and only upon those of the latter are the original definitions of intensity and extensity based. One could, of course, object that we are dealing with mere matters of definition. To this we would answer, *first*, that these definitions are not arbitrary, they formulate what actually happens in physical measurement; *second*, that the differences of actual procedure, upon which our analysis thus insists, are important matters of fact; and, *third*, that the very grasp of these operational differences has been hampered by the vague use of such terms as 'intensive,' 'extensive,' and 'additive.' It is even fair to say that understanding of the behavioristic frame of reference as a whole has suffered from this state of affairs. As was pointed out before (III 6) the failure to understand what is meant by the equality of units also contributed to the confusion.

4. If our analysis were to stop at this point, the reader would be justified in feeling that it suffers from a physicalistic bias. While it is correct that measurement in psychophysics *can* be considered simply as a regauging of physical dimensions by means of observer reactions, this is by no means the whole story of psychophysics. The psychologist is mainly, and properly so, interested in the *response aspects* of the phe-

nomena. To this phase our attention must now turn, and it will be best to begin with some very general remarks.

Like every other scientist the psychologist is interested in discovering the empirical laws (functional connections) that obtain between the variables of his field. Unlike the physical sciences, however, behavior science at its present stage finds that its variables are divided into groups, on the one hand the so-called response variables (R) and on the other, the group of experimentally manipulated variables (S). The functional relationships the behavior scientist is looking for are typically of the form

$$R = f(S). \tag{5}$$

While this is not the place to elaborate such a schema (3, 14), a few further remarks are necessary for the understanding of our last point concerning measurement and psychophysics in general.

(*a*) Neither S nor R is necessarily a single variable. They should rather be considered as representing groups of variables, so that S really stands for $S_1, S_2, \cdots S_n$, and R for $R_1, R_2, \cdots R_m$. The need for the introduction of such groups is particularly likely to arise on the side of the manipulated variables. Indeed, this is the comprehensible content of one of the main theses of the Gestalt writers.

(*b*) Both S and R are defined operationally by the scientist. The particular purpose of the investigation determines whether one single reflex (actone) or a relatively complex series of acts (action) is to be considered as one response variable and, likewise, which features of the physical environment are to be controlled or manipulated. That is to say, the choice of the descriptive units is entirely a matter of scientific expediency.

(*c*) S (stimulus!) always stands for a feature of the physical environment, R always for a behavior segment, but the psychologist is, by the nature of his material, forced to consider certain additional variables which fall into one of these two categories only *after they have been subjected to a certain amount of analysis*. These variables, let us call them

state variables, fall again into different groups, such as motivation (D), and training or previous experience (T). There are, moreover, the variables representing possible innate individual differences (I), which fall neither into the S nor into the R category. At the present stage of psychological knowledge equation (5) would thus have to be expanded into

$$R = f(S, T, D, I). \tag{6}$$

For our present purposes, however, equation (5) is sufficient, since it is characteristic of sensory discrimination (perception) that the response depends, within very broad limits, only upon clearly defined environmental variables (S). This feature could, indeed, be used in a behavioristic definition of what should reasonably be understood by perception. The extensive use of the term 'perception,' which the Gestalters advocate, makes it a weasel word that covers the whole of psychology only at the price of emptying it of any specific content. Historically speaking, this verbal preference is probably grounded in the intuitionistic philosophical background of the Gestalters.

(*d*) Many, and probably the theoretically most interesting, laws of contemporary psychology undoubtedly are of the forms (5) or (6) in which both stimulus and response variables occur. (The laws of physics would, in this terminology, appear as functional connections between S's only.) But there is also a body of psychological knowledge which consists primarily of response correlations that can be symbolized by either of the two formulæ

$$R_1 = f(R_2), \qquad F(R_1, R_2) = 0. \tag{7}$$

As a matter of fact any correlation between the scores on two different tests is of this type and so is much of our knowledge in the field of personality (correlation of traits). Most of the work of Lewin and his collaborators, if correctly analyzed, turns out to be concerned with $R - R$ laws and, since it lies in the field of human personality, freely uses the human yardstick (II 3). The theorizing of this group is largely a

rationalization of these two features (the interest in $R - R$ laws and the use of the human yardstick) which, in themselves, are easily understood consequences of their specific research interests. Historically speaking, these preferences, too, seem to have grown out of the atmosphere which the students of Kuelpe, Meinong, and in the last analysis, Brentano created at the German universities.[8]

As a matter of historical fact we are often able to discover a response correlation $R_2 = f(R_1)$, when we do not yet know how to ascertain the corresponding stimulus-response relationships

$$(8) \qquad R_1 = f_1(S, \cdots) \quad \text{and} \quad R_2 = f_2(S, \cdots).$$

If, however, $R_1 = f_1(S)$ and $R_2 = f_2(S)$ are both known, then it is always possible to compute the response correlation. In order to do that one has but to solve one of the two equations, say the first one, and then substitute the solution into the second. Symbolically this might be indicated in the following manner:

$$(9) \qquad R_2 = f_2(f_1^{-1}(R_1)).$$

5. The last remarks lead us back to the problem at hand, namely, the response aspects of the scales of psychophysics. Assume that two pitch scales have been constructed by two different procedures, one for instance by the method of constant stimuli, the other by fractionation. As has been said before, there is no *a priori* reason to expect that any two such scales should stand in a particularly simple relationship.[9] We notice that the relationship between any two such scales is just an instance of the general relationship (9).

Consider now the empirical law, recently confirmed by Stevens and Volkmann (18), that an observer, instructed to 'bisect' the interval between two given pitches P_1 and P_2 by

[8] Boring (5) has recently, with keen cultural sensitiveness, again called attention to these continuities.

[9] It can however be shown that any two scales obtained by the method of constant stimuli and the use of different percentage criteria are distinguished only by a multiplicative factor. This is a simple mathematical consequence of the way the j.n.d.'s are defined in this procedure (10).

adjusting a third tone so that it is 'halfway between P_1 and P_2 in pitch' selects a tone P so that

$$P - P_1 = P_2 - P. \tag{10}$$

where P, P_1 and $P_2(P_2 > P_1)$ are the numbers of a j.n.d. scale. Obviously this law, though it is of the type (7) could be expressed by any scale whatsoever, either in frequencies or by means of still another response scale. It is true, though, that if one uses the j.n.d. scale, the mathematical form of the law takes a particularly simple form. So far the case is completely analogous to that of our earlier example concerning the specific heat of substances (III 5).

Response scales, it is true, offer further advantages in the way of stimulating certain lines of research. In the case under consideration they direct the psychologist's attention towards the relationships between thresholds and supraliminal discriminations. The shape of the functional relationship between a physical and a response scale might also suggest certain lines of physiological research or, to use Boring's (4) happy expression, help the scientist to "tap" the organism at the right place. But such expectations are, of course, purely pragmatic and have no bearing on the methodological analysis of measurement.

Another interesting result of Stevens' experimentation (16) is that what has just been said about pitch does not hold for loudness. In other words, the numbers on a j.n.d. scale for loudness are not proportional to the numbers on a scale constructed by means of the bisection procedure. The point to be made is that it is neither helpful nor enlightening to describe this state of affairs by saying that one can construct an "additive" scale for pitch, while there is no loudness scale which enjoys such mensurational excellence. Such disguise of interesting results as a search for scales cannot possibly be to the advantage of psychophysics. Furthermore, it has the drawback that it tends to blur the really important distinctions of a behavioristic analysis of measurement. In the case under consideration, for instance, the so-called "additivity" of the pitch scale is based on the use of numerals by

the observer and this use is not clearly distinguished from the use of numbers by the scientist.

V. Summary and Conclusion

The results of the present discussion can be summarized into three points. *First*, it tries to discourage the use of such terms as 'intensive,' 'extensive,' and 'additive' with reference to scales in psychophysics. *Second*, it sets up measurement in psychophysics as a technique in its own right which one cannot without violence subsume under any of the customary classifications of physical measurement. Because of this methodological independence the term 'observer,' different from both 'judge' and 'subject,' has been consistently used throughout this section. So far as the response aspect of the scales is concerned, however, the observer acts really as a subject, while we have seen (IV 1) that as far as the rescaling aspect goes, his function is similar to that of a judge. *Third*, it points out the awkwardness of the fact that a field of research as significant as psychophysics has been disguised as a search for scales.

A concluding remark might help to forestall misunderstanding. No disparagement of the psychophysical scales is involved in their being set aside from physical measurement. Psychologists should indeed not rush into the use of physical terms if there is nothing else to be gained but a share in the prestige of the physical sciences. Behaviorism today is firmly enough established to make a more discriminating attitude completely compatible with our belief in the ultimate explanatory promise of the physiological level. The most fruitful expression of this belief at the present stage seems to us to emphasize the *theoretical* significance of the $S - R$ as opposed to the $R - R$ type of laws. On the other hand, it is methodologically perfectly conceivable to start the construction of science by measuring length and weight by means of observer-estimates, instead of with yardsticks and balances, provided only that the prescientific common sense level is once epistemologically secured. The actual excellence of physical measurement is entirely a matter of fact.

REFERENCES

1. BERGMANN, G. The logic of probability. *Amer. J. Physics*, 1942, 9, 263–272.
2. ——. An empiricist schema of the psychophysical problem. *Phil. Sci.*, 1942, 9, 72–91.
3. ——, & SPENCE, K. W. Operationism and theory in psychology. PSYCHOL. REV., 1941, 48, 1–14.
4. BORING, E. G. A psychological function is the relation of successive differentiations of events in the organism. PSYCHOL. REV., 1937, 44, 445–461.
5. ——. *Sensation and perception in the history of experimental psychology.* New York: Harcourt, Brace, 1942.
6. CAMPBELL, N. R. *Physics: the elements.* Cambridge: Cambridge University Press, 1920.
7. ——. *What is science?* London: Methuen, 1921.
8. CARNAP, R. *Physikalische Begriffsbildung.* Karlsruhe: B. Braun, 1926.
9. COHEN, R. M., & NAGEL, E. *An introduction to logic and scientific method.* New York: Harcourt, Brace, 1934.
10. EGAN, J. P. The measurement of sensory dimensions. Unpublished Master's Thesis, State Univ. Iowa, 1940.
11. MCGREGOR, D. Scientific measurement and psychology. PSYCHOL. REV., 1935, 42, 246–266.
12. NAGEL, E. Measurement. *Erkenntnis*, 1930, 2, 313–333.
13. RAMSPERGER, A. G. *Philosophies of science.* New York: Crofts, 1942.
14. SPENCE, K. W. Theoretical interpretations of learning. In *Comparative psychology* (ed. F. A. Moss; rev. ed.), New York: Prentice-Hall, 1942.
15. STEVENS, S. S. Volume and intensity of tones. *Amer. J. Psychol.*, 1934, 46, 397–408.
16. ——. A scale for the measurement of a psychological magnitude: loudness. PSYCHOL. REV., 1936, 43, 405–416.
17. ——, VOLKMANN, J., & NEWMAN, E. B. A scale for the measurement of the psychological magnitude pitch. *J. acoust. Soc. Amer.*, 1937, 8, 185–190.
18. ——, & VOLKMANN, J. The relation of pitch to frequency: a revised scale. *Amer. J. Psychol.*, 1940, 53, 329–353.
19. THOMAS, L. G. Mental tests as instruments of science. *Psychol. Monogr.*, 1942, 54, No. 3.
20. WOODWORTH, R. S. *Experimental psychology.* New York: Henry Holt, 1938.
21. *International Encyclopedia of Unified Science* (eds. O. Neurath, R. Carnap, C. W. Morris), Chicago: University of Chicago Press, 1938–.

REPETITION AND LEARNING—I. STIMULUS AND RESPONSE

BY THOMAS W. COOK

University of Toronto

I. Introduction

The once apparently secure position of repetition as a central factor in learning has of late become less secure. In fact, I know of no psychologist who holds unqualified allegiance to the frequency principle. The only effective repetition is, for Tolman, of the whole stimulating situation (8, 346); for Koffka, of occasion or process (4, 537). Thorndike, whose law of use was sacred script to associationists, has all but buried the law under an avalanche of experimental data. Yet the ghost still stalks through text-books on educational psychology. Is then the common sense view too firmly grounded in experience to be readily dislodged? Or has the concept of repetition been insufficiently analyzed? If the latter, an attempt to think through the problems involved might be worth while.

If as a first step we inquire into the meaning of the term, we find that it concerns the time dimension. In fact, the data from nonsense syllable experiments are often expressed indifferently in amount of time or number of repetitions. Now, time is not peculiar to psychology: It is studied, for example, by physics. When we look closely, however, we see a striking difference between the treatment of time in psychology and physics. Time to a physicist is merely a dimension within which certain processes occur. What occurs depends upon the nature of those processes and their relation to other processes. Mere duration as such has no significance. In contrast there is in psychological circles a notion that the occurrence of an event, quite apart from the nature of that event, has some intrinsic importance. We learn by repetition, we were told. How did such a conception arise?

The answer is partly theoretical and historical. It is a phase of the dominance of association theory and the synapse theory of the nervous system, and the use of nonsense syllable and conditioned response techniques in the laboratory. But there are other and less obvious factors.

(1) Many stimuli can be kept constant throughout a given learning process.

(2) Immediate memory.

All sorts of things are remembered after a single presentation and without special effort or techniques. Here mere occurrence in experience might plausibly be thought to effect the revival, and by analogy it is easy to pass to acceptance of repetition of experience as the essence of memorizing.

(3) A psychological process, in many instances, has an internal organization that marks it off from what goes before and what comes after. A repetition, even of a list of nonsense syllables, is a unitary act having a beginning and an end. Accordingly it is natural to stress the discontinuity as well as the similarity between successive phases of the learning process by speaking of 1, 2, 3, or more repetitions, and Humphrey's reduction of time to a fourth dimension of learning has seemed too strange to be readily absorbed into the current tradition (3).

Next, a distinction must be made between the occurrence of an event and the repetition of that event. This distinction might seem too obvious to mention, but indeed its application to psychological problems is far from simple. Even so keen an analyist as Peak confuses the two, and by so doing is able to pronounce repetition essential to learning without further examination of the problem. Now, no one would question Peak's statement that some events must precede learning (5). But learning involves change, by definition, and to call the first trial a repetition obscures what I consider the central issue: What features change and what remain constant? That once this question has been properly stated the answer is not obvious should meet with general agreement. Successive trials seem to be related by a wide range of qualitative and quantitative variations of similarity.

The gradual emergence of this point of view has in fact led to a loose and variable use of the concept of repetition. We apply the term when the similarity between one trial and the next is sufficient to pass a rough and ready inspection. But, however useful for practical purposes, this is a flimsy foundation upon which to build a scientific psychology of learning. Several basic principles must be accepted if we are to proceed.

(1) The problem of repetition involves discovering the kind and degree of identity between successive trials, and its effect upon the various phases and varieties of learning.

(2) The concept of partial identity must be substituted for that of similarity.[1] Such a reduction has long been advocated by association psychologists, and might be opposed by Gestaltists as the tenet of a rival school. But such a reaction would be undiscriminating. The purpose of the association school in equating similarity with partial identity was to reduce association by similarity to association by contiguity of parts. No doubt this tended to foster mental atomism, and to neglect of the problems of organization and response to relationships. But if we avoid atomism by granting that a part may be a relation or a whole quality, or even any organization or feature of an organized whole which can be identified, there is great advantage in reducing similarity to partial identity. In fact, all applications of mathematics to science are based on just this reduction. The concept of similarity is too indefinite for the formulation of scientific law, or even for a science based on classification alone. We must know what are the identical features in the situations concerned.

(3) There is no separate problem of transfer of training. Or, conversely, all learning (unless there exists a limiting case in which successive trials are identical on all counts) involves the problem posited in the transfer of training experiments: What identities and differences in successive trials affect what sorts of learning?

[1] For convenience I shall continue to use the terms similarity and partial identity interchangeably. But I shall *mean* partial identity throughout.

(4) Repetition implies discreteness as well as continuity. That is to say, a final solution of the problem of repetition must be related to rhythm and distribution of practice.

So much by way of introduction. To proceed some division of the field is necessary. I hope that the rationale of my division into stimuli and responses on the one hand, and the psychological processes of perception, skill, memory, thought, and motivation on the other, will be justified in the subsequent discussion.

II. Stimuli

Two definitions of stimuli are in current usage. The first of these, which I reject, defines stimuli in terms of (1) priority in time—S precedes R; (2) causation (or prediction)—S is the one necessary and invariable antecedent of R. Thorndike, who is perhaps the most active exponent of this translation of Hume into S-R terminology, cites 'thinking of a-b-c-d' as a stimulus and 'thinking of e' as a response. The historical and systematic arguments against the use of S to signify merely priority in time have been well stated by Robinson, himself an associationist (6, 39). But the view that S is the sole cause of R, I must also reject. The causes of psychological processes are to be found in the dynamic pattern of the situations in which they occur.

The second definition is given in Warren's *Dictionary of Psychology* as "an energy external to a receptor, which excites the receptor." This definition was developed by and is still essential to psychophysics; indeed, I cannot see how one can do adequate experimental work in the field of sensation and perception without adhering to it. In this field variations in the stimuli are stated in terms that are independent of the organism. In vision, the stimulus variables are wave length and wave amplitude, arrangement in physical space and physical time. And the brilliant success of psychophysics in working out the relations between stimuli, receptor processes, and sensation, makes defense of this restricted stimulus definition unnecessary.

When we come to the topics of learning and motivation, however, it is possible to ignore the energy-receptor definition

of stimuli.[2] We respond, we are told, to objects. And since the objects to which we respond can be more readily described in psychological than in physical terms, it is an easy step to calling any psychological process a stimulus, provided only that it is followed by some other process that we can call a response. The resulting looseness of terminology cannot be passed over if we are to make repetition a workable psychological term. We enter a house, let us say. Here house is S and the act of entering R. But what do we mean by house? A house as a physical object and a house as a percept belong in quite different systems of concepts, which have been developed to indicate aspects of reality connected only through intricate nervous processes. And, when we are behaving, the objects to which we respond are in the perceptual and not in the physical field. In Koffka's words, we respond to the behavioral and not the geographical environment.

To summarize: The energy-receptor definition of stimuli was formulated for experiments on sensation and perception, and was a tool in keeping distinct and working out the relations between the geographical (stimulus) and behavioral environment. Experiments in learning and motivation, however, have dealt chiefly with relations between perception on the one hand and memory or action on the other, and so have been able to neglect the stimulus-perception distinction.

But there is no reason to expect that learning any more than psychophysics can reach a satisfactory formulation of its processes or variables without exact definition of stimuli. My own conviction is that learning theorists and experimenters should adopt from psychophysics the working hypothesis that psychological processes are not closed systems, but are determined by processes in a world which can be and is described by physics in a set of concepts that are unambiguously objective, internally consistent, and independent of the organism.

As a matter of fact the present confusion arose, not so

[2] It is significant to find the division recurring in Gestalt psychology. Köhler and Koffka are interested in perception and emphasize the distinction between geographical and behavioral environment. Lewin's motivational approach mentions only person and environment.

much because of a prejudice against definition of stimuli in objective terms, but from the refusal to recognize that psychological processes *cannot* be so described. What is gained in clarity of thinking regarding the role of repetition in learning can be seen at once if the separation of geographical and behavioral environment is applied to a concrete learning task. A nonsense syllable as a stimulus does not change during learning, provided the external conditions are kept constant. But a nonsense syllable as an event in the behavioral environment (a percept, if you like) does change in ways which must be determined if the concept of repetition is to be retained in scientific psychology.

If, for the present, we confine the discussion to stimuli defined extra-psychologically, our task is then to find: (1) What temporal distributions of stimuli are invariable accompaniments of what kinds of learning? (2) Which stimuli, being absent, make learning less efficient or impossible? It will be realized that (1) does not imply that stimuli are the sole and sufficient causes of learning. Nor does it require that learning occur when any given factor is present. Some other essential may be lacking. We need only answer the question: Will learning proceed with equal ease and efficiency with or without the presentation of certain temporally extended stimulus patterns?

With these considerations in mind we may proceed to examine the limiting cases.

1. *No stimulus constancy.*—Can one learn when stimuli have no constancy? The usual answer would agree with that of Humphrey: "In an entirely chaotic world, where no two successive events were ever in any way similar, organic equilibrium, at least of the kind seen in repetitive learning, would be impossible for organisms of the type we know" (3, 275). One cannot improve one's response to chaos. It seems to me that there is no escape from this conclusion if we define stimulus constancy widely enough. Obviously learning *could* occur in a stimulus setting in which every aspect was constantly changing, provided the changes were in a constant direction. But Humphrey's statement would still hold true,

for constancy of direction of change means similarity between successive patterns.

2. *Completely constant stimuli.*—It is difficult to see how learning could occur without stimulus change. Experiments with nonsense syllables, based on the theory that we learn by repetition, aim to attain a high degree of stimulus constancy. The setting (lighting, position of subject, etc.) is carefully controlled, the experimental procedure standardized, and the lists uniform from trial to trial. Yet change, at least to the extent of the gradual removal of the sight of the nonsense syllables, is an important factor in the learning. And when we come to other types of learning, much less stimulus constancy is possible or desirable.

(*a*) Learning is conditioned upon the interruption as well as the recurrence of the stimulus. If the interruption is brief and regular we speak of rhythm; if considerable in extent we talk of distribution of practice. Both are stimulus factors listed in textbooks as favoring learning. On the other hand, effects such as fatigue, inattention, and inhibition accompany uninterrupted application of the stimuli.

(*b*) Stimulus constancy cannot be obtained in any learning situation in which variation in motor pattern is essential. Consider the relatively simple task of learning to throw darts at a target. Here there is some stimulus constancy; at least the target is there throughout. But there are at least two kinds of stimuli that are not constant: the light rays given off by the dart hitting the target and the kinæsthetic stimuli from the act of throwing. *And this inconstancy is essential to the learning.* For, as Thorndike has amply proved, perception of the relation between throwing posture and the place where the dart hits the target is necessary for improvement.

If we ask what degree of stimulus constancy and change are most favorable for such learning, we find outselves faced with a problem that has bedeviled the psychology of learning from its earliest origins. Perhaps an illustration from a different psychological field will serve to introduce the point I have in mind. I recall in graduate school one of the younger Gestaltists coming into the laboratory one afternoon and

announcing in great glee that he had discovered the perfect correlation—between the length of the shoe and the size of the foot. Now the bite of this illustration, for a mechanistic psychology, lies in the fact that any explanation which does not include purpose and design is simply and obviously silly. In learning, although the gradualness of the process makes their presence less obvious, the same factors are operative, and make it exceedingly difficult to predict from the character of S what R a subject will eventually make. All that one can say is that within certain limits the subject will adapt and fashion stimuli to his interests and purposes. It is accordingly impossible at present to describe or delimit the degree of stimulus constancy or change most favorable for problem-solving or acquisition of skill, except in terms of the requirements of the problem and the particular type of learning involved. In other words, we must defer an attempt at solution until we study the interrelations within the learning process itself.

We may, however, indicate one or two obvious limits to adaptability:

1. The subject's ability, native or acquired. Most of us make a rather elementary form of adaptation to the theory of relativity.

2. Certain stimulus features seem by their intrinsic nature to demand some definite type of response.

(*a*) Perception, as defined in elementary texts, conforms more closely to the characteristics of the physical world than a knowledge of local stimulation would lead one to expect. Gestalt and sensation psychologies, though they differ in that the former attribute more of this conformity to the pattern of the proximal stimulus and to the laws of organization, and accordingly less to learning, agree that certain stimulus properties are crucial for perceptual development.

(*b*) Action must shape itself to the nature of physical objects. One cannot readily learn to ride a horse without taking account of the horse.

(*c*) Certain intense stimuli have relatively constant relations to motivation; they seem to facilitate behavior leading

to their continuance or removal. Experience may improve one's technique of effecting that continuance or removal, but rarely leads to permanent loss of the tendency for identical stimuli to evoke the same general types of goal-directed response. The usual end result of a succession of pin-pricks is negative adaptation. But an inflamed tooth, though it may be unnoticed while one is getting out of a burning building, can ordinarily be expected to evoke actions quite similar to those previously called out by identical stimuli.

III. Response

The word response I am using in the strict behavioristic sense, to signify movements of muscles and secretion of glands. These have three aspects:

1. As stimuli. As physical processes acting on the sense organs, changes in muscular activity have been more or less adequately dealt with in the preceding section.

2. As components of psychological processes. Viewed as occurrences in the behavioral environment, movements of muscles belong in organized psychological wholes. This conception has been made familiar by Gestalt theorists.

3. As movements objectively observable, measurable, and describable without regard to functions (1) and (2). This last-named rôle of movement we will deal with briefly. Extensive discussion would overlap too greatly with other sections of the paper, entire omission would leave several points unclear.

(*a*) Exact repetition of movement in successive trials is exceedingly rare and difficult to obtain. For example, photographic reproductions of subjects learning ball-tossing show not only great variation from subject to subject, but also from trial to trial. Moreover, Snoddy's mirror tracing records show that later movements need not be selected from earlier ones. Pattern changes are the rule and not the exception (7). Even in the C-R set-up Culler has shown that if the situation permits, not only the CR but the UR alters until some optimum form is reached (1).

It is, in fact, only possible to exclude gross qualitative

variations in motor pattern from learning under one or other of three sets of conditions. (1) The animal is artificially restrained. This is the original Pavlovian set-up. The quantitatively measurable salivary response is alone permitted to vary. Even so there is some motor change in the early stages of learning, but this largely disappears in a short time and is neglected in Pavlovian theorizing. (2) The subject (an adult human) voluntarily limits himself to the required quantitative variation. (3) Through gradual adjustment the animal has reached a motor pattern that is optimal for the given set of conditions. In Culler's experiments the dog only lifts his paw quickly, effectively, and accurately in later trials.

(*b*) The changes in movement are not random, but adaptive in character. The limits of this principle and the problem involved have been considered in an earlier section.

(*c*) Changes in movement pattern as such (except indirectly through effecting perceptual changes) are only favorable to learning within fairly narrow limits. One cannot learn to skate by practicing pole-vaulting. If one looks at the data from both learning and transfer experiments, it would seem on the whole not unfair to conclude that some experience with a set-up in which one makes fairly similar movements is one essential to learning any motor task.

Such a statement is vague in the extreme and serves merely to introduce the problem, but for more exact qualitative or quantitative formulation we have little data. Indeed, as far as I know the problem has not previously been formulated as a study of all significant similarities between successive trials. How otherwise could we explain the general apathy toward the results of cross-education experiments, for whose positive transfer (running as high as 90 per cent) no satisfactory explanation has been offered?

Even Watsonian behaviorists, though restricted by their basic postulates to the observation of physical stimuli and muscular movements, have been slow to examine the actual identities between movements in successive trials, much less to discover which of these identities is significant for any particular type of learning. I am of course far from stating that

mere description of partial motor identities in successive trials is even an adequate description of repetitive factors in learning. But we shall at least avoid building theoretical structures on non-existent facts if we carefully observe actual identities and differences in this most accessible of all psychological fields.

IV. Stimulus-Response Relationships

Were this a prolegomena to a behavioristic psychology, rather than an attempt to formulate principles governing the application of the concept of repetition to learning, I should attempt to find the units upon which to build that psychology, not in S and R taken separately, but in S-R relationships. For not only do we know to what situations the subject is sensitive only through his behavior, but muscular movements are not psychological processes unless they occur in relation to the stimulus field. The study of the constant and changing features of psychological processes would therefore be study of S-R relationships, never of S or R in isolation. Accordingly, if S and R are defined as I have defined them throughout this paper, they are extrapsychological for behavioristic as well as subjective psychology. So far, then, we have not been dealing with psychological processes, but only with the study of physical stimuli and muscular movements defined in extrapsychological terms, as a preliminary to the study of the effects of these upon the change and constancy of psychological processes.

I cannot here defend the thesis (though I think it will bear critical examination) that progress in behavioristic psychology has resulted from just this separation of the description and control of S and R as physical variables, from that of the formulation and description of S-R relationships as psychological processes. Even in this statement the latter part is misleading if taken literally, for the physical S's and R's are not the same things as the S's and R's in S-R relationships. The definition of the latter, as well as their constancy and change during learning, belong to their properties as psychological wholes. That fact, I believe, is the central point of

Tolman's method of system-building, as well as (I speak with diffidence) the significance of operationism for a behavioristic psychology.

REFERENCES

1. Culler, E. Recent advances in some concepts of conditioning. Psychol. Rev., 1938, 45, 135.
2. Guthrie, E. *The psychology of learning.* New York: Harper, 1935.
3. Humphrey, G. *The nature of learning.* London: Kegan Paul, 1933.
4. Koffka, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace., 1935.
5. Peak, H. Negative practice and theories of learning. Psychol. Rev., 1941, 48, 319.
6. Robinson, E. S. *Association theory today.* New York: Century, 1932.
7. Snoddy, G. An experimental analysis of a case of trial and error learning in the human subject. *Psychol. Monogr.*, 1920, 28, no. 124, 47.
8. Tolman, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

THE SCREEN TEST IN MILITARY SELECTION[1]

BY WILLIAM A. HUNT, CECIL L. WITTSON, AND HERBERT I. HARRIS

United States Naval Training Station, Newport, Rhode Island

A 'screen test' in the common use of the words is an objective test, usually of the paper-and-pencil sort, given to a group of subjects in order to pick certain individuals to be eliminated or 'screened' from the group. As contrasted with tests for intelligence, screen tests are usually tests for aspects of personality such as neurotic tendency, emotional stability, etc. In military selection the screen test is used for weeding out the undesirables who would not be fit for military service. It is usually considered as supplanting the psychiatric interview. The main arguments in its favor are that it saves time and labor, and provides a more objective estimate. With the successful development of tests for intelligence, vocational aptitudes, personality traits, etc., it was inevitable that tests for the militarily unfit should be attempted. Unfortunately, the idea of such tests has developed under the full protection of the success and present authority of the other types of psychological test, and the value and probable success of screen tests in military selection has been largely assumed, and seldom called to critical account. The present article is an attempt at a theoretical evaluation of the screen test in military selection. The evaluation will consider two aspects: the theoretical suppositions behind the screen test; and the economics involved, the practical factors of cost, necessary facilities, ease of administration, etc., which are so often overlooked by the theoretician. Because the value of any practical instrument is relative, depending upon the other instruments available for the same use, the screen test will be compared to the brief psychiatric interview, its chief rival in military selection today.

[1] The opinions or assertions contained herein are the private ones of the writers and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

In picking out individuals who are undesirable for military service, any screening process, test, or clinical interview must accomplish three things. First, it must pick out the positive cases which show some neuropsychiatric handicap. Second, as a corollary of this, and just as important, it must not miss other positive cases. The criterion is not merely the number of positive cases selected by a test, but the ratio of these to the positive cases that are missed. Too often if a test is successful in picking out a large number of positive cases there is a tendency to evaluate the test by the absolute number selected for elimination without consideration of the number of cases which may have been overlooked. Finally, the screen test must avoid selecting false positives. By 'false positive' we mean a normal individual who is selected by the test as abnormal. This selection results in the screening out of individuals who are able to render good military service. Under other conditions than those of military selection, where the judgment rendered by the screen test is not vital to the economy of those using the test or to the lives and mental hygiene of those people being screened, the inclusion of a few false positives is not a vital matter. Thus, in using screen tests to select students for further mental hygiene treatment in a college community there is no harm done in including a few subjects for treatment who do not need it; nor is the false positive particularly harmful when one has available a large reservoir of man power. Thus, in military selection in the early days of the national emergency or in such peace-time business situations as selecting good insurance salesmen, it is doubtful whether the mistaken elimination of some good material is very vital. Today, however, with war economy straining our man power reserves to their utmost, the false positive is a vital matter. No one can estimate exactly the expense of such a loss, but it must be great. The usual criticism of military selection is made in terms of the number of undesirables that are left in the Service. It would be well to include in any critical estimate the mistaken elimination of men suitable for duty.

The philosophy of the screen test is definitely and soundly

empirical. It assumes that the undesirable individuals to be screened out are undesirable because of certain deviations in physique or behavior, and that testing situations can be devised in which these deviations will manifest themselves. The development of such testing situations and their validation is carried out experimentally in the best empirical tradition. The recent history of psychological testing contains many examples of the possibility of at least partially realizing this aim.

Screen tests for military selection, however, represent a more ambitious development of this philosophy, inasmuch as they usually represent a situation in which a test is called upon to measure not merely one characteristic or type of behavior but a number of them. Here the theoretical assumption runs somewhat as follows: If we can develop specific tests for each undesirable quality, we may then combine all these various individual tests into one general test for military unfitness. While theoretically sound, this shot-gun procedure runs into specific technical difficulties. The combining of a number of tests into one single unit may well result in a test which is too long to be practical in selection, or else in a test which is less long but unfortunately has a seriously lowered reliability. Moreover, any test item exists within the meaningful context of the other items which are given with it, and this dependence introduces a factor of relativity. What happens to a test item when it is included in a single test which combines items of an extremely diverse nature (a situation typical of screen testing for military selection) is an as yet relatively unknown factor. The further diligent application of item analysis techniques may possibly render the success of such screening procedures more probable than would appear at present. Future experimentation holds the only answer.

The most apparent obstacle to the successful development of military screen tests would appear to be the very multiplicity and diversity of neuropsychiatric disorders which must be screened out as contributing to military unfitness. Certainly any general screen test has to combine a number of

extremely different situations if it is to cover the entire gamut of neuropsychiatric disabilities, running from such neurological difficulties as multiple sclerosis and encephalitis through epilepsy, the post-traumatic personality disorders, homosexuality, emotional instability, neurosis, psychoneurosis, and the frank psychoses. The size of this task would seem to preclude the development of a single test and necessitate the substitution of a number of individual tests. As we go from a single test to a number of tests, however, the accent changes subtly from that of 'a screen test' to 'screen testing' which seems to be a modification of the ambitious hope for combining all screening in a single, unitary procedure.

The theoretical basis of the psychiatric interview is also empirical. Abnormalities occur and can be recognized. Through actual experience with clinical material the psychiatrist can learn to recognize these deviations. The question may be raised as to whether the interview has as sound an empirical basis as the test, since the criteria for judgment are not standardized as objectively or applied as rigidly in clinical practice as they are in test procedures. Objectivity and standardization are the big and vital contributions which the test can make. On the other hand, inclusiveness and flexibility come as advantages of the clinical interview. As we have said before, the very diversity of material stresses the necessity for breadth. No test as yet devised has the wide range made possible by the clinical experience of one well-trained psychiatrist. Moreover, the abnormality in question must be considered in the context of the total personality, and tests are not flexible enough for this at the present time. Progress is being made through the stress on 'patterns,' but this leads to multiple scoring (and the use of multiple weights and a series of stencils) and increases tremendously the time and effort involved, still without approaching the flexibility of the human mind. The very objectivity and standardization that we value in tests at the same time contribute an unfortunate rigidity.

This last argument may be summarized by saying that a test cannot *think*, and a psychiatrist can. On the other hand,

if the ability to think is to be counted as an advantage for the psychiatrist, he must use this ability to maintain the advantage; and integrative and creative thought will avail him nothing if training and clinical experience have not provided him with the material to use in his thinking. At the present time, no screen test has been developed to the point where it can function as efficiently as a competent, conscientious clinician, although some of those tests in use today certainly are preferable to an incompetent, careless clinician.

The requirements for screening procedures may be expressed as an arithmetical formula which we may call the 'Selection Index' (index of the excellence of the screening procedure). If we use the formula $\frac{P}{P + Pm + Pf}$, where P represents the positive cases who are unsuitable for military service, Pm represents the positive cases which are missed, and Pf equals the false positives, we have a rough formula for expressing the efficiency of such a test. If the test uncovers all the positive cases without picking any false positives, the index would stand at one, or perfection. As the number of false positives and misses increases, the value of the index will drop below one. It is impossible to suggest absolute standards along the continum furnished by such a ratio, as value still remains a relative matter. Thus, if one has available only an inefficient test, it may still prove to be better than none. As the development of screen tests proceeds, however, it should be possible by evaluating several different tests in a fixed situation gradually to evolve fixed points on this continum which might be labeled as good, bad, and indifferent. Such a task lies in the future, and at present the index is more valuable as a symbolic statement of requirements of a screen test than as an exact mathematical measure.

In addition to these general requirements, we must consider the specific factors which we have labeled 'economic.' The economics of screen testing is more important than is realized. The academic psychologist living in an environment where space, time and equipment are usually abundant, and where the official attitude toward psychological practices

is apt to be one of kindly indulgence, may easily forget that in the military situation it is necessary to strip every procedure to its essentials and even then to justify carefully these essentials in the eyes of authority. The problem cannot be met by merely assuming that the military attitude is prejudiced and wrong. One may run into prejudice, but in general one finds that the stress on economic factors is born of necessity, and there is no reason to assume that an applied psychology should do other than adapt itself to the task required of it. If screen testing or psychological testing in general is going to succeed in the military situation, it must be willing to submit itself to a realistic evaluation on economic grounds. In evaluating the economic factors we shall deal with them under arbitrary categories which we will call mechanics, man power, and time.

In discussing the economic factors involved in the mechanics of testing, the first thing to be treated is the nature of the tests themselves. It must be remembered that these economic factors, as is true of all the economic factors involved in military screening, are a function of the specific screening procedures used. A general discussion of them must necessarily be abstract and somewhat superficial. Tests available at present entail such varied impedimenta as a series of cards to be sorted, ink blots to be interpreted, and the question booklets of the various paper-and-pencil tests. The simplest tests, of course, are the paper-and-pencil tests, as they demand merely the test sheet or booklet itself and a pencil. If one is dealing with large groups of individuals who must be handled rapidly, it is obvious that a test which involves the manipulation of many complicated materials is out of the question, particularly since most of the tests of this sort are individual rather than group tests. Even the paper-and-pencil test has a serious complication inasmuch as it demands a writing surface and a room of some sort to be devoted to testing purposes. These everyday necessities of the academic world are not always easy to come by in a military situation. Moreover, paper-and-pencil tests vary widely in their complexity. The subject may make some sort of mark on the examination

booklet, or he may be given an examination booklet and a separate answer sheet. This last procedure involves a performance that may be exceedingly difficult for individuals of borderline intelligence and little academic experience. We have found at Newport that if one substitutes for the simpler test in which the individual marks the test booklet directly a question booklet-answer sheet technique, it definitely increases the difficulty of the task.

In comparison with the psychiatric interview the psychological test comes out a poor second as far as mechanics are concerned. The psychiatric interview demands no equipment except a small private room with a chair and table for the examiner. While even this equipment is sometimes difficult to obtain in a military situation, it is much more easily obtained than the larger room with desks, etc., which the paper-and-pencil test situation demands. If we consider only the mechanics of the two situations, we must conclude that the psychiatric interview is by far the preferable procedure.

The next factor to consider is that of man power, for both the psychological test and the psychiatric interview must be given by someone. In the case of the psychological test it may be administered by untrained people. Provided one has at hand personnel who are conscientious and trustworthy, a few hours training will suffice to instruct them in the administering and proctoring of paper-and-pencil tests. Here the test has a distinct advantage over the psychiatric interview, which can be given only by highly trained personnel. Moreover, it has been our experience that a psychiatrist must have a distinct flare for this work, as there are many excellent psychiatrists who cannot function efficiently at the pace and under the tension demanded by the rapid interview technique. To function at its best, psychiatric interviewing needs the best personnel available. Without question, the economics of man power give the psychological test a distinct advantage over the psychiatric interview, an advantage which may well be crucial in some circumstances.

The time consumed by any test is difficult to estimate beforehand since it varies with so many unforeseen circum-

stances. If one allows one hour as an average time for the paper-and-pencil screening test, given to groups of several hundred men at a time, the time per man is actually very short. The mechanics of scoring with its attendant bookkeeping, however, will add to this time. It also has been our experience that if the group to be tested at one time is too large the usual errors in administration are multiplied geometrically rather than arithmetically. The rapid psychiatric interview as used at the Newport Naval Training Station averages about three minutes, and 15 to 25 men may be handled in an hour. The literature contains several statements regarding the limited number of men that one psychiatrist can interview in a day. Fifty is the figure that has been set as optimum by the Surgeon General of the Army. Our experience at Newport would indicate that this is unnecessarily low. We have seen no indication that a psychiatrist cannot screen as many as one hundred men a day, and even more for short periods without diminution of efficiency. Unless one selects a specific test for evaluation, it is hard to give specific weight to the time factor. The wisest conclusion would be that the time factor probably favors the psychological test, although not so heavily as one might assume.

Whatever procedure is used in the preliminary screening, if some later form of recheck and more thorough study is not utilized, a large number of the individuals selected as unfit will be false positives, namely, men who would really be of value if retained in the military service. The Navy procedure as utilized at Newport provides for an observation ward. All men suspected during the preliminary screening examination are referred to this ward for further study before a decision is made to recommend separation from the Service. During the course of the man's residence on the ward he is seen by at least one other psychiatrist or psychologist who checks the original findings. Where necessary, various psychometric and laboratory procedures are also employed. Thus, a feebleminded man who is detected during the preliminary screening examination is not separated from the Service until he has been retested with both a group intelligence test and an

individual test. Electroencephalograms are obtained on the suspected cases of epilepsy, post-traumatic disorder, etc. If there is any doubt about the findings the final judgment is passed by two or more staff members in consultation. As a result of this procedure, we find that about one-third of the men who would have been discharged at the preliminary screening examination, are finally judged fit for Service and sent to duty.

Our discussion of the psychological test and the psychiatric interview as rival screening procedures in the selection of military recruits shows that both are based on the empirical method. The psychological test has the advantage of further empirical refinement in that it is standardized and objective. On the other hand, the psychiatric interview has the advantages of range and flexibility. If one considers the economic factors involved, it appears that the psychological test conserves man power and probably saves time. The interview, however, is much simpler in its mechanics. In view of the widely divergent conditions under which military screening takes place, it is impossible arbitrarily to select either method as the better. In each instance the judgment of 'better' can be rendered only after careful consideration of all the factors involved in each specific situation. In one instance the conservation of man power might well be the deciding factor which would lead to the selection of the test procedure. In another, the simplicity of the psychiatric interview might well result in this being used. If good psychiatrists are available, the advantages of range, flexibility and simplicity favor the interview. If the interviewing personnel is poorly trained or inadequate the factors of standardization, objectivity, time, and conservation of man power favor the use of the psychological test. At the present time, we feel that a good psychiatrist is a better screening instrument than a good test, but a good test is better than a poor psychiatrist.

Such evaluations have thus far been biased, inasmuch as the usual criterion against which the psychological test is measured is the psychiatric interview. Thus, we evaluate one

measure against the other. The final check will come when both methods are evaluated not against each other, but against their results in eliminating men who would break down under the stress of active military service. This means that the results of both methods must be checked through a careful follow-up of the subsequent records of the recruits involved. Such a procedure involves elaborate and arduous clerical work inasmuch as the service records of large numbers of men must be followed for long periods of time, a task which is not easily accomplished in war time. Until it is undertaken, however, the ultimate decision as to the success of both the psychiatric interview and the psychological test as screening procedures must be held in abeyance.

Up to this point, we have been dealing with the two methods individually. They are not, however, mutually exclusive and are often used together. Screen tests, for example, can be helpful to the psychiatrist in supplementing the results of his clinical interview. This is particularly true in the military situation where he is called upon to examine a large number of men in a brief period of time. It is our opinion that the best results in screening would be obtained if a psychiatric interview and a screen test were used to complement each other, since the examiner could thus utilize the potentialities of both.

THE NATURE OF THEORY CONSTRUCTION IN CONTEMPORARY PSYCHOLOGY [1]

BY KENNETH W. SPENCE

State University of Iowa

I. Introduction

The task of the scientist has been described as that of attempting to discover ever more generalized laws by which the observable events within his field of study may be brought into interrelation with one another. To this end he develops and refines (mainly in the direction of quantitative representation) his concepts or variables, arranges highly controlled (experimental) conditions of observation and introduces theoretical constructions. While it is not the primary purpose of this paper to attempt a methodological analysis of these components of scientific method, it is necessary to begin our discussion by calling attention to two somewhat different roles or functions that one of them, construction of theory, plays in different fields of science or in the same field at different stages of development.

In some areas of knowledge, for example present day physics, theories serve primarily to bring into functional connection with one another empirical laws which prior to their formulation had been isolated realms of knowledge. The physicist is able to isolate, experimentally, elementary situations, *i.e.*, situations in which there are a limited number of variables, and thus finds it possible to infer or discover descriptive, low-order laws. Theory comes into play for the physicist when he attempts to formulate more abstract principles which will bring these low-order laws into relationship with one another. Examples of such comprehensive theories are Newton's principle of gravitation and the kinetic

[1] The writer is greatly indebted to Dr. Gustav Bergmann for reading the manuscript and making valuable suggestions.

theory of gases. The former provided a theoretical integration of such laws as Kepler's concerning planetary motions, Galileo's law of falling bodies, laws of the tides and so on. The kinetic theory has served to integrate the various laws relating certain properties of gases to other experimental variables.

In the less highly developed areas of knowledge, such as the behavior and social sciences, theory plays a somewhat different role. In these more complex fields the simplest experimental situation that can be arranged usually involves such a large number of variables that it is extremely difficult, if not impossible, to discover directly the empirical laws relating them. Theories are brought into play in such circumstances as a device to aid in the formulation of the laws. They consist primarily in the introduction or postulation of hypothetical constructs which help to bridge gaps between the experimental variables. Examples of such theoretical constructs are legion in psychology, *e.g.*, Tolman's 'demand,' Hull's 'excitatory potential,' Lewin's 'tension system,' and a host of other mentalistic and neurophysiologically-sounding concepts. It is the purpose of this paper to examine the attempts of psychologists to discover general laws of behavior, particularly the auxiliary theoretical devices they have employed in doing so.

II. Theoretical Constructs in Psychology

Like every other scientist, the psychologist is interested in establishing the interrelations within a set of experimental variables, *i. e.*, in discovering empirical laws At the present stage of development the variables (measurements) studied by the psychologist and between which he is attempting to find functional relations appear to fall into two main groups:

(1) *R*-variables: measurements of the behavior of organisms; attributes of simple response patterns (actones), complex achievements (actions) and generalized response characteristics (traits, abilities, etc.). These are sometimes referred to as the dependent variables.

(2) *S*-variables: measurements of physical and social

environmental factors and conditions (present and past) under which the responses of organisms occur. These are sometimes referred to as the independent, manipulable variables.

While not all laws are quantitative, science typically strives to quantify its constructs and to state their interrelations in terms of numerical laws. The numerical laws the psychologist seeks may be represented as follows:

$$R = f(S).$$

The problem here is two-fold: (1) to discover what the relevant S variables are, and (2) to ascertain the nature of the functional relations holding between the two groups of variables.

In general, two radically opposed positions have been taken by scientists, including psychologists, as to the best procedure to follow in solving this problem. On the one hand are those who propose the introduction of theoretical constructs as described above. On the other there are the more empirically minded persons who attempt to refrain from the use of such inferred constructs and try to confine themselves entirely to observable data. An excellent defense of this latter viewpoint, along with a constructive proposal as to how such an approach can hope to discover general quantitative laws in psychology, is contained in the recent presidential address of Woodrow to the American Psychological Association (**23**). We shall leave consideration of the method proposed by Woodrow until later; certain criticisms he offers of the theoretical approach provide an excellent introduction to this method of discovering laws.

Beginning with the conception, more or less the same as that expressed at the start of this paper, that explanation in science consists in nothing more than a statement of established relationships of dependency (for psychology in terms of laws between measurements of environment and behavior) Woodrow goes on to protest that most psychologists seem to have been entirely too interested in postulating intermediate events occurring within the organism to explain the obtained measurements. The difficulty with such speculative con-

structs, he thinks, is that they cannot be measured because it is not possible to observe the interior of organisms. The result is that their specification must be left to the imagination. And as he says:

> . . . our imaginations have not failed us. The things we have stuck within the organism in the hope thereby of explaining behavior are almost without limit in number and variety. They include mental sets and cortical sets, traces, residues, synaptic resistances, inhibitory and excitatory substances, inhibitory and excitatory tendencies, determining tendencies, mental attitudes, sentiments, wishes, tensions, field forces, valences, urges, abilities, instincts, and so on and on. Very popular indeed is the animistic type of explanation (23, p. 3).

While it must be admitted with Woodrow that many of the theoretical constructs employed by psychologists have never been too satisfactorily specified, one must protest the lumping together of all theoretical constructs in such a completely indiscriminate manner. As a matter of fact, Woodrow has included in his list certain conceptions which were never meant to be explanatory concepts. Thus such terms as set, attitude, sentiment, and in some instances drive, are what Carnap (3) has termed *dispositional predicates or concepts*, because they refer to the disposition of an object to a certain behavior under certain conditions. They usually serve as names for events which do not appear in observable experience but instead are introduced into the scientist's language in terms of conditions and results which can be described in terms that refer directly to observable experiences. Such concepts are prevalent in all fields of science and serve a useful purpose.

Then, again, Woodrow has failed to distinguish in his list between what turn out upon analysis to be very different kinds of theoretical constructs. While some of them are little better than the animistic notions of primitive man, others have qualified as quite satisfactory in the sense that they have led to the formulation of behavioral laws. We turn now to the consideration of the different kinds of theories (theoretical constructs) that have been proposed in psychology.

III. Four Types of Theoretical Constructs

Theoretical constructs are introduced, as we have said, in the form of guesses as to what variables other than the ones under control of the experimenter are determining the response. The relation of such inferred constructs (I_a) to the experimental variables, measurements of S and R, is shown in the following figure. Here we have assumed an over-simplified situation for purposes of exposition.

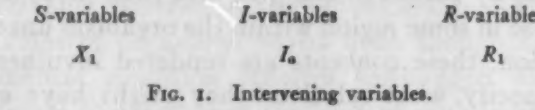

Fig. 1. Intervening variables.

If under environmental conditions X_1 the response measure R_1 is always the same (within the error of measurement) then we have no need of theory. Knowing that condition X_1 existed we could always predict the response. Likewise if, with systematic variation of the X variable, we find a simple functional relation holding between the X values and the corresponding R values we again would have no problem, for we could precisely state the law relating them. But unfortunately things are not usually so simple as this, particularly in psychology. On a second occasion of the presentation of condition X_1, the subject is very likely to exhibit a different magnitude of response, or in the second example there may be no simple curve discernible between the two sets of experimental values. It is at this point that hypothetical constructs are introduced and the response variable is said to be determined, in part by X_1, and in part by some additional factor, or factors, I_a, $I_b \cdots$, *i.e.*, $R = f(X_1, I_a, I_b \cdots)$. The manner in which these theoretical constructs have been defined by different psychologists permits a grouping of them into four categories: (1) animistic-like theories in which the relations of the construct to the empirical variables are left entirely unspecified, (2) neurophysiological theories, (3) theories involving constructs defined primarily in terms of the R variables and (4) theories involving constructs intervening between the S and R variables.

1. *Animistic conceptions.*—Little need be said about such instances of psychological speculation. They are included here merely for the purpose of completing the record. The invoking of such general concepts as the 'soul,' 'mind,' 'élan vital,' 'entelechy,' 'idea,' 'libido,' not to mention many more specific instances (*e.g.*, insight, instinct [2]) in order to account for the apparent capriciousness of the behavior of organisms has been all too prevalent in psychology. When not safe from disproof by reason of the fact that their locus is usually specified to be in some region within the organism unaccessible to observation, these concepts are rendered invulnerable by failure to specify what relations they might have either to the *S* or *R* variables. While such vagueness renders them unverifiable, it does insure them a vigorous and long career among certain types of thinkers. Needless to say, such vague conceptions receive little attention today among scientific-minded psychologists.

2. *Neurophysiological theories.*—The extent to which neurophysiological concepts, defined in terms of the operations and instruments of the neurophysiologist, are employed in psychological theorizing is not nearly so great as is sometimes thought. As a matter of fact, if we employ such variables to help us out in our formulation of behavior laws we are not, strictly speaking, theorizing for such concepts are not hypothetical, but are empirically defined. In such instances we have stated a law interrelating environmental, organic and behavioral variables. As yet we do not have very many such laws, except in the case of the simplest kinds of behavior (sensory responses, reflexes, etc.).

There are, of course, many theoretical constructs in psychology which are supposed to represent hypothetical neurophysiological processes, but whose properties are defined either in terms of the response variables (type 3 theory), in terms of environmental factors and the response variables (type 4 theory), or just assumed to be operating without making any specification of their relations to either the

[2] Such terms, of course, when used as dispositional predicates serve the useful function of providing a name for the phenomenon.

environmental or response variables (type 1 theory). Examples of these are Köhler's construct of brain field (12) to explain perceptual and memory phenomena (type 3), Pavlov's constructs (16) of excitatory and inhibitory states (type 4), and certain neural trace theories of learning (type 1).

It will be seen that this category really cuts across the other three. Further consideration of some of these theories will be given in our discussion of the final two classes of theory.

3. *Response inferred theoretical constructs.*—The fact that the behavior of organisms varies even though the objective environmental condition remains unchanged has led some psychologists to assert that the laws of such behavior cannot be formulated in terms of objective environmental variables even though additional hypothetical constructs are employed to bolster the effort. These writers have insisted that behavior must be accounted for in terms of the psychological situation. Thus Lewin (13) in his book *Dynamic theory of personality* refers to what he describes as the complete failure of such German writers as Loeb, Bethe and other objectivists to develop an adequate theoretical interpretation of behavior in terms of the objective situation, *i.e.*, the physical situation as described by the operations of measurement of the physicist and/or the objective social situation as described by the sociologist. It is always necessary, he insists, to describe the situation as the subject sees or perceives it, *i.e.*, in terms of what it means to him. Typical quotations from Lewin's writings indicate the positive tone taken by such writers:

> For the investigation of dynamic problems we are forced to start from the psychologically real environment of the child (13, p. 74). Of course, in the description of the child's psychological environment one may not take as a basis the immediately objective social forces and relations as the sociologist or jurist, for example, would list them. One must rather describe the social facts as they affect the particular individual concerned (13, p. 75).
>
> One of the basic characteristics of field theory in psychology, as I see it, is the demand that the field which influences an individual should be described not in 'objective physicalistic' terms, but in the way in which it exists for that person at that time (15, p. 217).

As Lewin implies in the last quoted excerpt it would seem that this type of 'psychological' approach to the theoretical constructs of psychology is characteristic of the self-styled field theorists or Gestalt psychologists. Thus Koffka (10) makes use of the construct of 'behavioral environment' and the more inclusive construct of 'psychophysical field' which includes the former and the physiological field, while Köhler (11, 12) refers to 'phenomenal field' and to 'brain field.' Koffka and Köhler differ slightly from Lewin in that they introduce a physiological terminology in the description of the properties of some of their behavior-determining fields whereas Lewin does not. The methods of determining the structure and properties of these fields whether 'brain field,' 'behavioral environment' or 'life space' are, however, es-

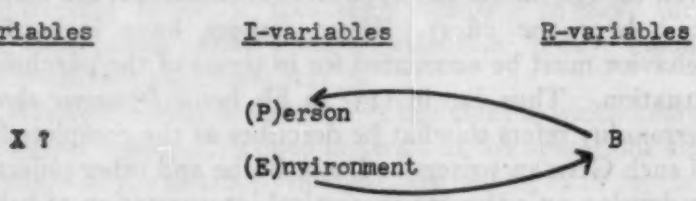

FIG. 2. Lewin's theoretical constructs.

sentially the same, and as we shall see later, involve extensive use of the phenomenological type of introspection.

The nature of this type of theorizing may be made clearer by attempting to show how it fits into the schema we have already employed. Figure 2 makes use of the constructs of Lewin, who has been the most articulate of this group of writers so far as the exposition of the formal nature of his theorizing is concerned. Lewin employs the concept of life space to represent the totality of facts which determine the behavior (B) of an individual at a certain moment. The life space includes two groups of constructs, the person (P) and the psychological environment (E). Use is then made of certain concepts from geometry (topology and hodology) and dynamics to represent the existing relationships.[8] By

[8] There is considerable reason to doubt whether Lewin does much more than take over the terms of topology, making little if any use of the postulates (implicit

means of what Lewin calls coordinating definitions these constructs are said to be related to empirical concepts.

Without going into detail, Figure 2 reveals an interesting fact. It is that little, if any, use is made of the *S*-variables in Lewin's theorizing.[4] The question immediately arises then as to what kinds of laws, if any, does Lewin arrive at. The functional relationship which frequently appears in his writings, $B = f(P, E)$, is obviously not a law of the type that psychologists were said to be interested in, *i.e.*, $R = f(S)$. But if it is not this kind of a law, what kind is it? Some writers have implied that Lewin really does not attain any laws at all. They call attention to the fact that laws are statements of relations between *independently defined variables* and they ask what are the two sets of independent variables in Lewin's formula.

The answer to this question is not easy, and the writer is of the opinion that it has not been met in all instances of field theory. However, an examination of the methods employed by Lewin and his students in determining the structure and properties of their fields reveals that they depend heavily upon the phenomenological introspections of their subjects or themselves. If we now think of these as a kind of verbal response or 'perceptual' response, in which the subject tries to describe his own particular way of perceiving the objective situation, we see that Lewin's theory really does provide us with laws mediating between independent variables, *i.e.*,

definitions) of this formal system. Koch, after making a very thorough analysis of Lewin's formulations, states that Lewin "finds it expedient to abstract from the postulates (of topology) the properties with which they implicitly endow the constructs, instead of fully stating the postulates" (9, p. 148). In this sense, then, it may be said that Lewin employs very fragmentary parts or sub-systems of topology. As for the formal system of dynamics it remains thus far closeted, to use Koch's happy description, in Lewin's mind. Like so many of these field theorists, Lewin sets up a most attractive program for theory. Taken in conjunction with his interesting experiments the illusion is nicely created that there is some connection between them.

[4] Lewin states that the objective physical and social surroundings "have a relation to the life space similar to that which 'boundary conditions' have to a dynamic system" (15, p. 217). He believes that physicalistic behaviorism has made the mistake of treating such variables as if they were parts of the life space. Attention should be called here to the fact that Koffka concerned himself much more extensively with the objective environmental variables or the geographic environment as he called it. He discussed at some length the relations between the geographic and the behavioral environments—the traditional problem of perception (10).

between two different responses of the subject, or as in some cases, between the experimenter's own perceptual responses and the subjects' subsequent response. Thus Lewin discovers what amount to laws of the following type: $R_1 = f(R_2)$.

Of course, such theorists do not always rely on such phenomenological introspections. Once certain laws of the above type have been formulated it is possible to formulate further laws between purely overt behavior items, neither of which are of this introspective, verbal type. There are also instances in which from the pattern of the observed response the theorist makes an inference as to the nature of the hypothetical field; and then by means of his postulates as to what happens in these fields he is able to make certain predictions as to subsequent behavior. An excellent example of the latter is Köhler's theoretical treatment (12) of perceptual problems involving reversible figures and the effects of prolonged inspection in certain types of simple perceptual situations. Thus in the light of perceptual behavior to reversible figures, he formulated the hypothesis that percept processes are associated with fields of electric currents in the nervous system. Then by means of postulates based on knowledge about electrolytical conduction he was able to predict other perceptual behavior.

By and large, however, the field theorist depends heavily upon phenomenological introspection in introducing his theoretical constructs. In order to understand the field of the subject he asks him to describe how he perceives the situation, or he infers it on the basis of his own introspections. With engaging frankness Snygg has made an appeal for the recognition of the important role that phenomenological introspection plays in these theories.[5] Thus he writes in connection with the problem of prediction:

> By postulate *B* the determining locus of action is the behaver's p.f. [phenomenological field]. This is not open to direct observation by an outside observer. The process of prediction therefore involves

[5] Snygg also admits another obvious characteristic of such theoretical systems which some of its proponents have not always willingly acknowledged. Reference is made here to the fact that such systems are anthropomorphic.

two steps: (1) the securing of an understanding of the subject's field by inference or reconstruction, (2) the projection of the future field.

The first operation is of the common "Now why did he do that?" or "Under what circumstances would I have done that" character. Much of the topological work of Lewin is of this type and essentially the same procedure was used by Shepard when from the behavior of his rate he inferred the existence of floor cues which he himself was unable to experience (17, p. 413).

That this field approach to the problems of psychology has been fruitful and valuable is amply supported by the experimental contributions it has made, although in the writer's opinion, the theoretical superstructure has played a much less significant role than is sometimes credited to it. Furthermore, the phenomenological approach has its advantages, particularly in the complex field of social behavior of the human adult. It is obviously much easier to gain some notion as to the relevant variables determining such complex behavior by asking the individual to verbalize than it is to employ the procedure of trying to hypothesize them from knowledge of past history. Usually the latter is not available in sufficient completeness to make it even worthwhile to try to theorize as to the nature of such historical laws.[6]

A final point of no little importance is the failure of such field theories to provide us with laws which will enable us to control and manipulate the behavior-determining psychological field. Such laws are obviously a basic prerequisite to successful clinical therapy. While it may be true, as Snygg claims (17), that psychiatrists and teachers find the phenomenological approach most valuable in diagnosing behavior disorders, it is difficult to understand how the response-response laws it provides can be of much use in guiding therapeutic treatment. The latter requires a knowledge of what to do to the individual, what changes in his physical and social environment to arrange, in order to bring about the desired behavior changes. The laws telling us how to

[6] The situation in the case of animal behavior is somewhat different. Here one usually does have a pretty good record of the past history relevant to the present environmental situation.

proceed in such matters are historical laws and involve as an important component of them objective variables representative of past and present factors in the physical and social environments. Psychiatrists and clinical psychologists who employed a purely phenomenological approach might or might not be successful at diagnosis; it is difficult to see how they could ever prescribe satisfactory reëducative procedures.[7,8]

4. *Theoretical constructs as intervening variables between S and R variables.*—In sharp contrast to these response inferred theories with their emphasis upon the phenomenological approach is the point of view that theoretical constructs in psychology are to be regarded as 'intervening variables' which bring into relation with one another the dependent *R* variables on the one hand and the independent *S* variables on the other. As Bergmann and Spence (**2**) have previously pointed out, two psychologists in particular, Hull and Tolman, have advocated, each in his own individual way, this type of psychological theory. In a little known paper Tolman (**22**) has presented an excellent account of such a theoretical program, while in his new book *Principles of behavior* Hull (**7**) has demonstrated in actual practice how such intervening variables provide us with a formulation of the basic principles or laws governing simple learning behavior. The following discussion makes no attempt to give a systematic account of this theoretical procedure. Instead we shall merely outline very sketchily its main features and then single out one or two aspects of it for more detailed examination.

[7] Bergmann has summed up this difficulty most succinctly in the form of the following questions:—"But even so, what is the predictive value of the suggestive metaphor 'psychological environment'? Is it not the business of science to ascertain which objective factors in the past and present states of the organism and its environment account for the difference in response, so that we can actually predict it instead of attributing it, merely descriptively and after it has happened, to a difference in the psychological environment?" (**1**).

[8] Mention should perhaps be made here of the fact that the theoretical constructs (factors) that Spearman (**18**), Thurstone (**20**), and other factor analysts arrive at are response derived and hence fall into this class. These men do not, of course, use the phenomenological method, but beginning with response intercorrelations (empirical $R - R$ relations) they arrive at hypothetical factors by various methods of mathematical analysis. Like the phenomenologists their theoretical factors have no tie-up with the *S* variables.

According to Hull and Tolman, theoretical constructs, or intervening variables have to be introduced into psychology either when we do not know all the important variables entering into a set of experimental events, or the precise nature of the interrelating function is not known. Consider, for example, the data obtained from conditioning experiments. These investigations have presented us with a wealth of data showing how the response variable changes or varies with the manipulation of certain other experimental variables. That is to say, various measurable aspects of response are studied as functions of the manipulable environmental variables and the data so obtained are plotted in the form of various curves.

The task of the psychologist here is to discover the precise nature of the interrelations holding within this set of variables. Instead of knowing merely that the response, R, is some function of the variables X_1, X_2, X_3, $\cdots X_n$, he desires to know the precise function. But in such a situation, involving as it does a large number of variables, the function relating the dependent and independent variables is so complicated that we are unable to conceive of it directly. It is necessary, say Hull and Tolman, to proceed by conceiving of it as broken down into successive sets of simpler component functions. These component functions begin by introducing new intervening constructs defined in terms of the independent variables. Further intervening variables are then introduced by stating them as functions of the first set of intervening constructs, until finally the dependent behavior variable is postulated to be a function of one or more of the intervening variables.

Thus Tolman, beginning with the empirical data that the response measure is some function (f_1) of two groups of independent variables (environmental variables and individual difference variables), writes:

> In place of the original f_1 function, I have introduced a set of intervening variables, I_a, I_b, I_c, etc., few or many, according to the particular theory. And I have conceived a set of f_2 functions to connect these intervening variables severally to the independent variables on the one hand, and an f_3 function to combine them

together and connect them to the final dependent variable on the other (21, p. 9).

It is characteristic of Tolman's theorizing, however, that it never gets beyond the programmatic stage. In his writings Tolman has merely shown how such a theoretical device as the 'intervening variable' can provide for the definition and proper utilization within psychology of such mentalistic terms as 'demands,' 'hypotheses,' 'traits,' 'discriminanda,' etc., but he never actually reaches the point of formulating a specific theory. In the present context this would, of course, require the precise specification of the various functions relating the intervening variables to the independent and dependent experimental variables. Instead of risking guesses on such matters, however, Tolman seems to prefer to ascertain them empirically by a series of what he calls 'standard experimental set-ups.' He believes the data from these studies will mirror the functions obtaining between the experimental (empirical) and intervening (theoretical) variables.[9]

Quite in contrast to such an approach, Hull has ventured to make guesses as to the precise nature of the functions introducing the intervening variables in his theoretical formulations. Thus he has attempted to formulate the basic laws of simple adaptive behavior (learning) by introducing a number of intervening variables.[10] Beginning with the experimental variables, he has introduced by means of specific mathematical functions such symbolic constructs as stimulus trace (s), habit strength ($_sH_R$), the limit of habit strength

[9] Tolman has been accused (and he has usually made no denial) of employing the phenomenological method in his psychology, and, because he has worked with animals, of being guilty of anthropomorphism. The present writer's interpretation is that it is Tolman, the experimentalist, who uses phenomenological introspection; Tolman, the theorist, introduces his intervening variables in terms of objectively defined variables. The difference between Lewin and Tolman on this point is interesting. Lewin, as we have seen, employs the phenomenological method primarily in his theoretical efforts, whereas Tolman uses it chiefly in the formulation of experimental problems.

[10] Reference is made here to Hull's latest writings in which the 'intervening variable' technique is made more explicit. As Bergmann and Spence (2) have pointed out, Hull's earlier miniature systems (5, 8) really involved the definition of such mediating constructs.

(M), excitatory potential ($_sE_R$), inhibitory potential (I_R), effective excitatory potential ($_s\bar{E}_R$), and so on. Ultimately the observable response variable, R, is stated to be some function of the final intervening variable (*e.g.*, $R = f(_s\bar{E}_R)$). Despite the neurophysiological tone of some of the terms that Hull employs to designate these constructs, the mistake should not be made of interpreting them as physiological concepts. Their scientific meaning is given only by the equations introducing them, and in this respect they are strictly comparable to many similar, abstract, mathematical constructs employed by the physicist in his theorizing. The use of neurophysiological terms and such additional statements as Hull sometimes makes as to their possible locus in the nervous system merely serve the purpose of providing experimental hints to persons interested in such matters. It may or may not turn out that they represent actual neurophysiological states or conditions that will some day be measurable by independent neurophysiological procedures.

An example of the specific manner in which Hull introduces his theoretical constructs is shown by the equations which he employs to define the two constructs, habit strength ($_sH_R$) and the limit of habit strength (M). With all experimental variables except the number of reinforced trials (N) and the length of the delay of the goal reinforcement (L) constant, the two equations are:

$$_sH_R = M(1 - e^{-iN})$$
$$M = 100e^{-kL}.$$

Grice (4) has recently shown how such precisely defined theoretical constructs may be tested. He employed several mazes of different absolute lengths involving a shorter and longer path to the goal and ran different groups of rats on each maze. On the basis of the above two equations the following rational equation was then derived mathematically to describe the rate of learning the mazes:

$$\left[N = b \log \left(\frac{e^{-kL} - e^{-kHL}}{e^{-kL} - e^{-kHL} - a} \right)\right].$$

Where N = number of pairs of trials on the two paths to learn the maze.
L = length of short path to goal.
H = ratio of long to short path length.
k, a, b = empirical constants.

This rational equation was then shown to fit the experimental data, whereas another equation

$$\left[N = b \log \left(\frac{\log H}{\log H - C}\right)\right]$$

derived from a logarithmic postulate [11] as to the relation of M to L was shown not to be in agreement with the experimental data.

Until constructs are introduced in some such precise fashion as Hull employs one really does not have a scientific theory, for it is only under such conditions that the possibility of verification or refutation exists. Unfortunately, much of what has passed for theory in psychology has been sadly lacking in this respect, a state of affairs which is largely responsible for many of the 'theoretical' controversies, and for the low regard in which theory is held in some quarters in psychology. That theory construction has not always been intelligently pursued, however, is no reason for doing without theory. Without the generalizations which theories aim to provide we should never be in a position to predict behavior, for knowledge of particular events does not provide us with a basis for prediction when the situation differs in the least degree. The higher the level of abstraction that can be obtained the greater will be both the understanding and actual control achieved.

IV. The Ultra-Positivistic Approach

All the methods of ascertaining the laws of psychology we have discussed so far have agreed, in principle at least, that it is necessary to introduce some type of symbolic construct. It is also apparent that agreement ceases as regards the

[11] $M = 100 - K \log L$.

extent to which the proponents of these different views have insisted on rigorous and objective specification. We turn now to a quite different approach to the same problem—that of the ultra-positivist or empiricist, who tries to eschew all types of theoretical constructs. Usually the writings of such persons are limited to negativistic, critical attacks on all theory. Recently, however, Woodrow has come forward with a constructive proposal as to how general mathematical laws of psychology may be discovered by a method which he believes avoids the necessity of introducing theoretical constructs.

Woodrow's method consists in an attempt to obtain by mathematical curve fitting a general equation describing a wide variety of experimental facts. Thus, after plotting a series of experimental curves of such widely varying situations as learning to abstract, learning to associate numbers and letters, learning a maze, reaction time to different intensities of stimulation, the forgetting of monosyllabic words, brightness and pitch discrimination, the growth of intelligence, etc., Woodrow sought to fit these empirical curves by means of a single general equation. He found that such an equation could be found and that it took the following form:

$$Y = a + \sqrt{p^2 + k^2(1 - f^{x+d})^2}.$$

This equation states a law between two experimental variables, a dependent response variable Y, *e.g.*, errors, successes, latencies, etc., and an independent manipulable variable X, *e.g.*, number of practice periods, intensity of the stimulus, preparatory interval, etc. But it will be noticed further that the law includes more than these two variables. It also involves certain constants or unknowns, termed parameters, the a, p, k, f, and d in the equation. We cannot stop to discuss these parameters in too great detail here. Suffice it to say that the specific shapes of the different empirical curves determine what parameters it is necessary to assume. Two of them, a and b, have no particular psychological significance, they merely express the fact that either one or both variables may have been measured by scales

with an arbitrary zero. The parameter, k, is introduced because all his curves exhibit a limit to improvement, no matter how favorable the status given the environmental variable. Another parameter, f, is determined by the rate of approach of the curve to this limit and p, finally, is introduced to take care of the fact that the lower part of the curve sometimes shows positive acceleration.

As Woodrow himself points out, these parameters may be thought of, if one so wishes, as representing hypothetical states or factors within the organism.[12] Woodrow prefers not to do so, for as he argues, it really makes little difference what the internal referents are since they cannot at present be independently measured anyway. From the point of view of finding a general equation or law that will fit the experimental data the important thing, Woodrow states, is to determine how many parameters are required and the mathematical function of each.

While in general sympathy with Woodrow's mathematical approach and his view that it is unnecessary to specify the factors or complexes of factors inside the organism which determine the values of the parameters, the writer is, nevertheless, of the opinion that such an equation as Woodrow obtains by his analysis is, on the whole, rather barren and sterile. Its defect is not that the factors *within* the subject are not specified, but rather that it fails to give any indication whatever of the conditions or variables even *outside* the subject which determine these parameters. In this respect Woodrow's approach is similar to the field theorists'. We shall have occasion later to point out other resemblances between these two approaches.

This criticism can be made clearer, perhaps, by contrasting the end result of Woodrow's empirical procedure with Hull's rational approach to the same problem. Woodrow's law

[12] Woodrow writes, "Now these parameters may refer to anything whatsoever, conscious, physiological, environmental, psychic, or purely imaginary. Here one is free to follow his predilections, whether for motives, excitatory and inhibitory substances, field forces, states of disequilibrium, inertia of the nervous system, abilities, or what not" (23, p. 4).

specifies but a single experimental variable determining the response:

$$Y = f(X_1).$$

Hull's theorizing culminates in a much more comprehensive law. Thus in the case of his theoretical formulation of simple adaptive behavior (learning) his derivation involves the following series of steps:

$$(1)\ \boxed{M} = f(T, G)$$
$$(2)\ \boxed{H} = f(\boxed{M}, T', N)$$
$$(3)\ \boxed{D} = f(T'')$$
$$(4)\ \boxed{E} = f(\boxed{D}, \boxed{H})$$
$$(5)\ \boxed{I} = f(N, W, F)$$
$$(6)\ \boxed{\bar{E}} = f(\boxed{E}, \boxed{I})$$
$$(7)\ R = f(\boxed{\bar{E}}).$$

Here the squared symbols are intervening variables or hypothetical constructs. The other symbols represent the dependent response measure (R) and the various manipulable, environmental variables (T, G, T', etc.). By substituting in the successive equations, a single equation stating R as a function of seven environmental variables is obtained.[13]

$$R = f(T, G, T', N, T'', W, F)$$

The latter procedure thus comes much closer to achieving the goal of the scientist, that of discovering all of the experimental variables determining the response measure and the nature of the functional interrelations holding between them. If this is achieved, the parameters become known functions of these experimental variables and thus become experi-

[13] The reader may ask: Why have a series of equations that introduce intervening variables? Why not write the single equation from the beginning and avoid the hypothetical constructs? One obvious reason, of course, is that it is just not possible to conceive of such a complex function all at once. As Tolman says, one can arrive at it only by breaking it down into a series of simpler functions. The reader is referred to a recent article by Hull in which he gives other reasons for using intervening variables with multiple equations rather than a single equation (6).

mentally manipulable. Woodrow's formulation, on the other hand, provides us with very little more information than we had when we started.

It is also interesting to note that a strong case can be made out for the position that Woodrow's method is really not a great deal different from those theoretical approaches which infer their constructs from the characteristics of the response. In introducing his parameters Woodrow is, in effect, assuming or postulating some kind of hypothetical factor. Thus, on noticing that some of his curves show an initial period of positive acceleration, Woodrow assumes a factor, p, "whose influence is greatest when the magnitude of the environmental variable is small" (23, p. 7). This factor is inferred, we see, from the characteristics of the response curve and is therefore in a certain sense akin to the hypothetical constructs of the field theorists which, as we have seen, are also inferred from the response characteristics. The important difference is that in arriving at these hypothetical factors Woodrow does not make use of the introspective report associated with a response but rather bases his constructs on the mathematical properties of a curve of successive response measures.

V. Conclusions

In summary, the present paper has stated the task of the psychologist to be that of discovering the general laws of behavior, and has attempted to present a brief and critical outline of five different methods of approaching this task. The conclusions that the writer believes may be drawn from this survey are:

1. That theory is still at a very primitive level in psychology, concerning itself primarily with the discovery of low-order laws rather than the integration of different realms of laws.

2. That there is a variety of different theoretical procedures possible in psychology.

3. That some psychologists substitute, often quite unconsciously, phenomenological introspection and anthropo-

morphic thinking for theorizing. There is, of course, nothing wrong with such introspection; it has often served as a means of formulating interesting and valuable experiments. In such instances, however, the credit should not be given to a theory.

4. That many theories in psychology have provided us with response-response (*R-R*) laws rather than stimulus-response (*S-R*) laws.

5. That the most promising theoretical technique, especially from the point of view of discovering the historical stimulus-response laws, is the so-called 'intervening variable' method proposed by Hull and Tolman.

REFERENCES

1. Bergmann, G. Psychoanalysis and experimental psychology: A review from the standpoint of scientific empiricism. *Mind*, 1943, 52, 122–140.
2. ——, & Spence, K. W. Operationism and theory in psychology. Psychol. Rev., 1941, 48, 1–14.
3. Carnap, R. Testability and meaning. *Philos. Sci.*, 1936, 3, 419–471; 1937, 4, 1–40.
4. Grice, G. R. An experimental study of the gradient of reinforcement in maze learning. *J. exp. Psychol.*, 1942, 30, 475–489.
5. Hull, C. L. Mind, mechanism and adaptive behavior. Psychol. Rev., 1937, 44, 1–32.
6. ——. The problem of intervening variables in molar behavior theory. Psychol. Rev., 1943, 50, 273–291.
7. ——. *Principles of behavior*. New York: D. Appleton-Century Co., 1943.
8. ——, Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning*. New Haven: Yale University Press, 1940.
9. Koch, S. The logical character of the motivation concept. II. Psychol. Rev., 1941, 48, 127–154.
10. Koffka, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace & Co., 1935.
11. Köhler, W. *Gestalt psychology*. New York: Liveright Publishing Corp., 1929.
12. ——. *Dynamics in psychology*. New York: Liveright Publishing Corp., 1940.
13. Lewin, K. *A dynamic theory of personality*. (Trans. by D. K. Adams and K. E. Zener) New York: McGraw-Hill Book Co., 1935.
14. ——. *Principles of topological psychology*. (Trans. by Fritz and Grace Heider.) New York: McGraw-Hill Book Co., 1936.
15. ——. Field theory and learning, pp. 215–242. In *Forty-First Yearbook National Society for the Study of Education*, Part II. Bloomington, Illinois. Public School Publishing Co., 1942.
16. Pavlov, I. P. *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. (Trans. and ed. by F. C. Anrep.) London: Oxford University Press, 1927.
17. Snygg, D. The need for a phenomenological system of psychology. Psychol. Rev., 1941, 48, 404–424.
18. Spearman, C. *The abilities of man*. New York: Macmillan, 1927.

19. SPENCE, K. W. Theoretical interpretations of learning. In *Comparative Psychology*, rev. ed. (F. A. Moss, ed.), New York: Prentice-Hall. 1942, Chap. 11.
20. THURSTONE, L. L. *The vectors of mind.* Chicago: Univ. Chicago Press, 1935.
21. TOLMAN, E. C. The determiners of behavior at a choice point. PSYCHOL. REV., 1938, 45, 1–41.
22. ——. Operational behaviorism and current trends in psychology. *Proc. 25th Anniv. Celebration Inaug. Grad. Stud.*, Los Angeles. The University of Southern California, 1936, pp. 89–103.
23. WOODROW, H. The problem of general quantitative laws in psychology. *Psychol. Bull.*, 1942, 39, 1–27.

IF-THEN RELATIONS IN PARALOGICS

BY H. M. JOHNSON

Tulane University

Dr. Christenson's recent article (2) on the applicability of generalized if-then relations and on the possibility of reinstating certain fallacies ordinarily called equivocation, affirmation of the consequent, and denial of the antecedent, originated, undoubtedly, in a set of good motives. Nevertheless, the article is distinguished by certain defects which I am impelled to point out lest the future historians of science should accept it as a specimen of reasoning which passed critical inspection of the American psychologists in 1942.

I must begin by noting a distinction which some may consider to be rather finely drawn, but which impresses me as being fundamental. Dr. Christenson, in his title and also in his text, mentioned 'The if-then relation.' In exact discussion, such as one demands of a logician even though he may not expect it, the word *the* connotes *uniquity*. For example, when we speak of *the* now President of the United States, *the* second son of Jacob Brown, *the* Prince of Darkness of this World, etc., we make it plain that we are naming *classes*, each of which contains *at most one* member. But, if we mention *a* man of sorrows, *a* public-spirited citizen, *an* implacable opponent of the New Deal, it is evident that each descriptive phrase designates a class which *may* contain more than one member, inasmuch as its defining form does not constitute it as one-membered or singular.

Similarly, if we speak of *the* if-then relation, our description implies that there is *at most one* such relation. I believe that this manner of speaking is likely to obscure certain important distinctions. We may therefore profitably consider some generalities.

Given any relation R, there is a class of terms which it connects; this class is called the relation's *domain*, for it sets

them into a certain *order*, determined by its own properties, and therefore *dominates* them. Given, furthermore, two relations R_1, R_2, with corresponding domains D_1, D_2: if R_1 and R_2 have the same properties in respect to reflexiveness, symmetry, transitivity, etc., then at least some of the terms in D_1 stand in a one-to-one correspondence with at least some of the terms in D_2. These two sub-classes of terms, being dominated by *similar* relations, are therefore said to be *isomorphic* (*cf.* 4).

But although R_1, R_2 are alike in respect to those characteristics which we have mentioned, we should not assume that they are therefore *identical*. For, either of them may be intrinsically capable of dominating one class of terms and no others. For example, the relation i of material implication dominates a class of propositions. It can dominate nothing but propositions. The relation c which I once called 'eventual causality' (3) is like i, in that it is reflexive, non-symmetrical, transitive, many-many. Its domain is isomorphic with a part of the domain of the relation i. But while i dominates a class of propositions and nothing else, c dominates a class of events, and nothing else. Is there not, therefore, some difference between $R_1 = i$ and $R_2 = c$, which prevents either from invading the domain of the other? I think so, and therefore I consider c to be different from i, although they are alike in respect to all the traits which I have specified.

Now there are many relations which, like i and c, are reflexive, non-symmetrical, transitive, and many-many. Each of them may properly be called *an* if-then relation, but this is not a good reason for treating them as if they were identical. Again, the relation of strict implication I is, like the relation of material implication i, reflexive, non-symmetrical, transitive, many-many. Also it is like i in that it dominates propositions and nothing but propositions. But it is unlike i in that it does not dominate a part of the class of terms which lie within the domain of i. Both are if-then relations, but they are not identical. Similarly we may compare the relation c of eventual causality with the relation C which I called 'intrinsic causality' (3), and notice corresponding similarities

and differences. If we treat all if-then relations as if they were identical we thereby equivocate.

Dr. Christenson exhibits *an* if-then relation which he denotes by the symbol $\rightarrow$ and defines by a table of truth-values. He calls it a relation of 'implication'; and in terms of truth-value definition, it is identical with the relation of material implication i. For, if p, q are propositions,

$$p\, i\, q\cdot = \cdot(pq')',^{1}$$
$$p \rightarrow q\cdot = \cdot(pq')'$$

the upper equation being a formal definition and the lower derived from Dr. Christenson's truth table. Hence, in this context, and by the axiom of equals,

$$\rightarrow\cdot = \cdot i.$$

In a manuscript which the author kindly permitted me to read, he formally asserts this equivalence, and also says that the relation $\rightarrow$ is 'weaker than' the relation i or the relation c. Both assertions cannot be true: Hence it seems to me that the author should revise and clarify his meaning.

Of course, if the author insists on regarding the relation $\rightarrow$ as being capable of dominating a class of events, as well as the propositions that assert them, then his *truth-table* can be transformed into an *occurrence-table*. However, as I have suggested above, I believe that similar relations should be distinguished by the fields which they are capable of dominating.[2]

Before we can entertain certain questions of primary importance to scientific method, we need to scrutinize Dr. Christenson's truth-table (2, p. 486) and notice the relations which he defines by means of the truth-values of various combinations of symbols. Some of the terms which he thus defines do not agree with interpretations put on them in the text, and these discrepancies have to be mentioned.

[1] In this notation the symbol ′ negates the expression to which it is suffixed.

[2] The confusion which I have just now indicated characterizes Boring's book (1), in which *correspondence*, *correlation*, *isomorphism*, are severally equated to *identity* and therefore to each other.

First of all, the relation which he denotes by the symbol $\rightarrow$ and calls 'implication' has the same set of defining truth-values as the relation called material implication in a two-valued logic. Second, the relation which he denotes by the symbol $\succ\!\!-$ and calls 'entails' or 'presupposes' must be read "P presupposes Q," or "Q is presupposed by P," not "Q presupposes P." For it is readily shown that in terms of truth-value definition, the relation 'presupposes' is interchangeable with the relation 'materially implies' which Dr. Christenson denotes by the symbol $\rightarrow$. For, by definition, p and q being propositions,

$$p\,i\,q\cdot = \cdot p \rightarrow q\cdot = \cdot(pq')'$$

i.e., it is false that q is false and p true. Thus, if $p \rightarrow q$, the truth of p depends on the truth of q while the truth of q is independent of the truth of p. Therefore we say that while the truth of p is *sufficient for* the truth of q, nevertheless the truth of q is *necessary to* the truth of p. For, since p cannot be true if any of its implicates such as q be false, therefore the truth of q is necessary to the truth of p, or in other words, the truth of p presupposes the truth of its implicate q.

Dr. Christenson seems to wish to read $Q \succ\!\!- P$ as equivalent to "Q presupposes P." This reading, however, contradicts his table of truth-values; moreover, it contradicts the assertion on page 487 of his text, that the relation $\succ\!\!-$ is *converse* to the relation $\rightarrow$.[3]

Given two relations R, R', each of which is converse to the other, if t_1Rt_2, then $t_2R't_1$, and *the same truth values hold for both relationships*. The rule is to replace the asserted relation by its converse, and interchange the related terms.

According to Dr. Christenson's truth-table the relation

[3] Perhaps the author has used the word *converse* in some sense that is foreign to logic. If so, he ought not to have done it. I submit two descriptions of mutually converse relations.

"If R is any relation, the *converse* of R is the relation which holds between y and x whenever R holds between x and y. Thus *greater* is the converse of *less*, *before* of *after*, *cause* of *effect*, *husband* of *wife*, etc. (5, *i*, p. 32)."

"Two relations, R and R', are each other's converse if they are such that, whenever one of them holds between the terms of a couple (t_1, t_2), the other holds between the terms of the *inverse* couple (t_2, t_1). . . . It is sufficiently obvious that every relation has a converse (4, p. 166)."

$\succ\!\!-$ is *not* converse to the relation $\rightarrow$. The truth-table actually takes the interpretation that *presupposes* or *entails* is equivalent to *implies*, not converse to it.

The relation of double implication, which Dr. Christenson denotes by the symbol $\leftrightarrow$ has, in truth-table definition, exactly the same meaning as *is interchangeable with*, which some writers denote by the symbol $\equiv$ and others by the symbol $=$. Some authorities doubt whether the distinction between *equivalence* ($\equiv$) and identity ($=$) is worth preserving. This question does not greatly concern the present discussion: for, in truth-value definition $\leftrightarrow$ is interchangeable with the symbols ($\equiv$), ($=$) indifferently.

Dr. Christenson's truth-table therefore reduces itself to a form with which many readers are quite familiar: namely,

TABLE I

P	Q	$P \rightarrow Q \cdot = \cdot Q \succ\!\!- P$	$P \leftrightarrow Q \cdot = \cdot P \equiv Q$
T	T	T	T
T	F	F	F
F	T	T	F
F	F	T	T

Note, however, that $Q \succ\!\!- P$ must be read "Q is presupposed by P," "P presupposes Q," "Q is necessary to P," "Q' entails P'," or the like. In terms of truth-value definition, it *cannot* be read "Q presupposes P."

Dr. Christenson then argues for the respectability of such fallacies as equivocation, affirming the consequent, and denying the antecedent, as useful procedures in 'scientific' inference. Logically, of course, they are crimes or nuisances, just as they have always been. However, I believe that Dr. Christenson intends to point out a certain utility of certain classes of false hypotheses; if so, there is some merit in his argument. I shall set forth, as briefly as possible, some reasons why each of these fallacies *may* lead to good results, but in doing so I shall not try to expound Dr. Christenson's views.

Consider:

$$(p_1 \rightarrow q)(p_1 = p_2) \cdot \rightarrow \cdot p_2 \rightarrow q$$

in which all the terms are propositions. If the first relationship stated in the left member is true, then the truth of the conclusion is guaranteed if the assertion $(p_1 = p_2)$ is true. (Here we use the symbol = to mean that the two terms which it connects have the same truth-value.) But let us suppose that p_2 and p_1 are not identical, and also that their equivalence in the sense that we have just mentioned has not been proved or disproved. What, then, have we? A formal equivocation. But the whole proposition as we have stated it is *hypothetical;* the only means of testing the factual truth of the hypothesis $(p_1 = p_2)$ or of the conclusion $(p_2 \rightarrow q)$ is by *observation;* thus the test becomes an *experiment.* So long as we do not assert the equivalence without proof, no harm has been done. The hypothesis, which may be false, has catalyzed our curiosity into an experimental plan. But even so, the facts do not prove that any 'scientific' inference based on equivocation is formally valid.

Similarly with respect to the considerations mentioned in (2, p. 491 *f.*): If we rely on the method of induction we thereby rely on the validity of affirming the consequent. This procedure we know to be a fallacy. Its fallacious character is merely concealed not eliminated by the well known 'inductive leap,' for the leap itself is an instance of affirming the consequent. As a sample of 'scientific inference' the procedure is putrefactively soft and nothing can make it sound. But so long as we control our eagerness to set forth a conclusion whether it be valid or not, we can utilize the method of induction. We need not affirm the consequent at all: we can merely let the premises remind us of one or more *hypotheses* which imply the facts of observation; we can remind ourselves that hypotheses are not the results of inference; we can analyze them for their implicates, and test some of the implicates by observation.

The corresponding fallacy called denying the antecedent stands in similar case. But I see no need for lengthening this discussion.

In brief: No artful manipulation of symbols according to prescribed rules can make good logic out of bad logic, or

produce a proof of the validity of any procedures which are invalid by the postulates and rules of a genuine logic. The *structure* of science as we know it is pre-determined by the definitions, postulates, and rules of manipulation of symbols that we call modern logic. This logic includes the whole of the traditional or Aristotelian logic, cleared of certain well known defects; it includes a great deal that Aristotle and his imitators overlooked. But since the corrected Aristotelian logic is a 'proper part,' though a small part, of the logic of science, we may be sure that if any procedure assumes equivocation, affirming the consequent, denying the antecedent to be valid, then it does not yield a set of rules for 'scientific inference.' It belongs to the field of antilogic and *therefore* of anti-science.[4]

BIBLIOGRAPHY

1. Boring, E. G. *The physical dimensions of consciousness.* New York: Century Company, 1922.
2. Christenson, J. A., Jr. The if-then relation and scientific inference. Psychol. Rev., 1942, 49, 486–493.
3. Johnson, H. M. Rival principles of causal explanation in psychology. Psychol. Rev., 1939, 46, 493–516.
4. Keyser, C. J. *Mathematical philosophy.* New York: Dutton, 1922.
5. Whitehead, A. N., & Russell, B. *Principia mathematica.* Cambridge: University Press, 1925.

[4] If this were a question of personal rivalry—as most happily it is not—I should be taking advantage of Dr. Christenson's present status by offering this contribution now. He is doing his duty with the armed forces, without ready access to certain manuscripts and publications that he needs. However, *tempus fugit* for the aging civilian as well as for the valiant solder. Perhaps it is wise, if either of them has anything to do or say, to get it over while he still can.

BRIGHTNESS ENHANCEMENT IN FLICKERING LIGHT[1]

BY THEODORE LOUIS JAHN

State University of Iowa

In recent years our knowledge of the activities of visual sense cells has increased tremendously, and there has also been a rapidly increasing interest in the neural elements of the optic pathway. As a result of this latter interest it has been pointed out repeatedly (4, 7, 8) that many visual phenomena cannot be predicted merely from a knowledge of the visual mechanism of the sense cells, and that the form of the sensation produced is determined largely by the neural elements of the pathway. This is to be expected since all of the sense cells do not act as a single unit and since one should not assume that the complex neural elements of the visual pathway merely transmit sense cell impulses without modification. It is obvious, therefore, that the photochemical cycle of the sense cells should not be expected to explain all details of vision.

Whenever the controlling mechanism in any given phenomenon is ascribed to the activities of either sense cells or neural elements it is essential to know precisely why one of these explanations is to be preferred above the other. In the case of brightness enhancement in flickering light the explanation given by Bartley (3, 4, 7) for eliminating the sense cells as the controlling factor does not seem to be adequate. The present paper is a brief summary of the evidence for the point of view that the photochemical changes in the sense cells may be the causative mechanism of enhancement.

In recent publications Bartley (3, 4, 7) has pointed out that in human vision as the frequency of a flickering light is lowered to a point below that of fusion, the Talbot effect dis-

[1] Aided by a grant from the Rockefeller Foundation for work in cellular physiology.

appears and the apparent average brightness increases to a maximum which may be twice that of the same light when viewed continuously. This is the so-called 'Bartley effect.' A comparable phenomenon was discovered by Brücke (12) who noticed an enhancement of brightness of the flash (rather than of the average intensity as studied by Bartley) of light obtained from a rotating black and white disc. Bartley (3, 4, 7, 8) has repeatedly claimed that these phenomena cannot be explained on a photochemical basis. He states (7, p. 139):

> The discovery of enhancement . . . add[s] another phenomenon to the list of those which seem contrary to photochemical predictions, for nothing that has yet been said about the sense-cell would have led to a critical repetitive rate for enhancement of brightness over a level produced by steady illumination. One of two things must occur. Either conceptions of sense-cell behavior must be radically reformulated, or else the phenomenon must be attributed to the neural part of the optic pathway.

In line with the purpose of the present paper it should be pointed out (1) that enhancement occurs in photosynthesis where there is no possibility of a neural explanation, (2) that enhancement can be predicted from the steady state equations of the visual photochemical cycle in the light-adapted eye, (3) that the photochemical equation which Bartley (3) considered to be inadequate is apparently that for the dark-adapted eye and should not be used for explaining visual phenomena in the light-adapted eye, (4) that enhancement is a general property of all photochemical cycles, and (5) that the existence of a photochemical explanation does not disprove the neural theory.

Enhancement of photosynthesis.—An enhancement phenomenon which is strictly comparable to the Bartley effect has been known for some time to occur in the case of photosynthesis and was probably first recorded by Brown and Escombe (11). Warburg (24) demonstrated quite conclusively that in the photosynthetic activity of the *alga Chlorella* the effectiveness of a given quantity of light used in continuous illumination could be increased 100 per cent if the

light flickered at 133 cycles per second with a light to dark ratio of 1/1. Emerson and Arnold (16) demonstrated that the effectiveness of flickering light could be increased as much as 400 per cent over that of continuous illumination with a light to dark ratio of 1/5 at 50 cycles per second, and their results indicate that a further increase should be expected if the light to dark ratio were further decreased. This effect is directly comparable to the Bartley effect when the frequency of retinal stimulation is in the region of enhancement. At present there is no complete set of data on photosynthesis which is comparable in all respects to that of Bartley on brightness enhancement, but it seems quite probable that if the frequency of flash were increased sufficiently during photosynthesis the enhancement effect would disappear and would be replaced by a typical Talbot effect. The data of Emerson and Arnold indicate that the frequency for a Talbot effect on photosynthesis should be well above 400 cycles per second.

The relation of enhancement to the mechanism of the photosynthetic cycle has been discussed by several investigators (16, 1, 13, 10), and it is generally agreed that the higher effectiveness of intermittent light results from the changes in the average relative velocities of different phases of the photochemical cycle.

A chemical basis for visual enhancement.—The photochemical cycle of the eye is usually represented by the equation of Hecht (18) in which

$$S \underset{\text{dark}}{\overset{\text{light}}{\rightleftharpoons}} P + A,$$

where S is a photosensitive substance (*e.g.*, visual purple) and P and A are both decomposition products and precursors of S (*e.g.*, visual yellow). It is understood that one or more intermediate steps (*e.g.*, the formation of transient orange) may exist without appreciably altering the kinetics of the above reaction. One of the transient products of the decomposition of S (sometimes referred to as 'new' $P + A$ and

which may be transient orange) is supposed to catalyze a secondary reaction which results in stimulation of the nerve. The electrical manifestation of this secondary reaction is probably the electroretinogram (**22, 25**).

On the basis of available evidence the sequence of events seems to be as follows: The photosensitive substance (S) is converted into decomposition products ($P + A$) through a transient intermediary stage. This intermediary product catalyzes the production of an electrical change in the sense cell, and this electrical change stimulates the nerve.

Whenever the eye is exposed to continuous illumination an equilibrium is reached between S and $P + A$. Equations which describe the percentage of S as a function of intensity of the light have been formulated by Hecht and are discussed in several publications (**17, 18, 19, 20, 21**). It is apparent that by considering the necessary modifications introduced under conditions of flickering light an enhancement phenomenon could have been predicted from an inspection of these steady state equations.

In the case of flickering light a pseudo-steady state is produced. Furthermore the steady state equations describe only the *average* quantity of S. If the rate of flicker is very high the actual quantity at any moment will not vary much from the average. However, as the rate of flicker is decreased the actual quantity fluctuates with the same frequency as that of the flicker, and at low frequencies the magnitude of the fluctuations may become large. Since each flash begins when the quantity of S is at a peak of the fluctuation, the amount of S decomposed by the flash to new $P + A$ (and consequently the magnitude of the sensation produced) will be greater than that obtained from the same quantity of light under conditions of steady illumination at the same intensity. Another way of expressing this fact is to say that at the beginning of the flash the eye is less light-adapted than it would be in continuous light of the same intensity. As the frequency of a stimulus consisting of a given light to dark ratio is further decreased the apparent intensity *during each flash* should decrease because of light adaptation during the

flash, and this should give a decrease in the total average intensity as observed in the Bartley effect. An enhancement phenomenon which passes through a maximum can therefore be predicted from the photochemical cycle and the mere presence of an enhancement does not necessarily mean that neural factors are involved.

In view of certain statements by Bartley (3) it seems necessary to clarify some of the conditions under which the steady state equations for the equilibrium between S and $P + A$ and the equation for the thoroughly dark adapted eye are applicable. Bartley (3) stated:

> It is easy to understand how two opposing processes would produce a lower concentration of substance x [*i.e.*, $P + A$] than if the one producing it were acting alone, and thus give the Talbot effect. But it is not obvious how an increased concentration could thus be produced (regardless of the rate of alternation between light and darkness). It is not clear how the net result of the action of two opposing processes could exceed that of one process acting by itself, and thus produce the Brücke or Bartley effect.

These statements apparently apply to the thoroughly dark-adapted eye where the concentration of $P + A$ is very small and the amount of S converted to $P + A$ (*i.e.*, x) would be greater for constant than for flickering light. However, under conditions of light adaptation (and all flicker experiments are necessarily on partially light-adapted eyes) it is not the equation for the dark-adapted eye but rather the equations for the pseudo-steady state which are applicable. It has been pointed out above that steady state equations apply under conditions of a high flicker rate and that a modification of these equations is applicable to flickering light at low frequencies. In these latter cases the amount of $P + A$ produced by each flash is compared not with that formed in an eye with a very high concentration of S (in which there is 'one process acting by itself') but with that formed from an eye which is *continuously* illuminated and in which the concentration of S is at the level postulated by the steady state equation. In brief, no true equilibrium between S and $P + A$ is reached at low flicker rates, and the amount of

$P + A$ formed at each flash may be considerably larger than that formed by the same amount of light under a steady state equilibrium, and therefore the enhancement may occur. The objection of Bartley (3) apparently arises from the consideration of the wrong equation.

The inherent mechanism for enhancement in all photochemical cycles.—Photochemical cycles consist of several reactions of which only one is influenced by light. The other reactions occur equally well in light or dark. As a result of illumination a photosensitive substance is converted by the action of light into another substance which is unaffected by light. This second substance may undergo several intermediate chemical changes of which the net result is the formation of more of the original photosensitive substance. Either the second substance or one of the intermediate products is responsible for the production of a secondary reaction which results in photosynthesis, vision, or some other measurable manifestation.

During the dark period between flashes of a flickering light the amount of photosensitive substance accumulates to a concentration higher than that possible under continuous illumination of the same intensity. Since the effectiveness of a given quantity of light depends upon the concentration of the photosensitive substance, it is apparent that at some frequency which is determined by the velocity constants of the individual reactions the effectiveness of flickering light should be greater than that of a constant light of the same intensity. Therefore, the mechanism for enhancement is necessarily inherent in every photochemical cycle. The only condition under which enhancement should not be expected is when regeneration of the photosensitive substance is prevented. In that case, however, the normal cycle no longer exists, and the system must be considered as a chain of reactions of which at least one step is irreversible.

Photosynthesis and vision comprise two of the most important biological phenomena which have a basis in photochemistry, and huge amounts of data have been gathered about both processes. The principal similarity between them is that a photochemical reaction forms one link in a complex

chemical cycle, and that in each case the cycle, although widely different in most respects, permits an enhancement phenomenon to occur. Data which may be interpreted as evidence of enhancement in photochemical cycles are not limited to photosynthesis and human vision, but they have also been noted for vision of arthropods by Mast, Dolley and others (23, 15) and for photographic plates by Davis (14). Unfortunately, no apparent attempt has been made to link the arthropod data with the photochemical cycle, but there seems to be no reason for necessarily considering this effect for arthropods as being caused by the refractory period of the sense cells or of the nervous elements as proposed by Mast (23). Apparently these data could also be explained on a purely chemical basis.

Possibility of a neural explanation.—In spite of the existence of a photochemical explanation of enhancement there is a considerable amount of evidence that in human vision the effect may be neural. One of the best lines of evidence is that the flash rate for maximum sensory effect is supposed to be constant (8–10 per sec.) for all intensities studied and for all practical light to dark ratios. The data published by Bartley (3) are not complete enough to be entirely convincing on these points, but if his statements regarding the unpublished data are accepted, then these data might be considered good evidence of a neural effect. On the basis of a purely chemical effect one should expect the frequency of maximum enhancement to vary with both intensity and light to dark ratio. However, whether this variation would be great enough for experimental demonstration is a matter for further consideration. Another excellent item in favor of the neural theory is that under certain conditions facilitation in the cortex might be expected to vary with the alpha rhythm. This variation does occur in the rabbit (9, 2, 3) and apparently also in man (6). A further item in favor of the neural theory is that if the field surrounding the test object is illuminated the enhancement does not pass through a maximum at 8–10 cycles per second but continues to increase at the lowest frequencies studied (5). Therefore, the enhancement predict-

able on a photochemical basis is not necessarily the one discovered by Brücke and Bartley, and the possibility remains that these effects may be entirely neural.

It is certainly not the purpose of the present note to declare that brightness enhancement is purely a photochemical effect, but to point out that the reasons used by Bartley (3, 7) for considering the photochemical explanation unworkable are not valid. A misunderstanding of the reasons why one theory is to be preferred over another certainly does not assist in the advancement of knowledge, and it may lead to considerable confusion. It is quite possible that the correct explanation of brightness enhancement is not photochemical but neural. However, if the photochemical explanation is to be completely eliminated from consideration this must be done on a basis quite different from that which has been used.

BIBLIOGRAPHY

1. Arnold, W. Kinetics of photosynthesis in Chlorella. *Cold Spring Harb. Sympos. quant. Biol.*, 1935, 3, 124–127.
2. Bartley, S. H. Temporal and spatial summation of extrinsic impulses with the intrinsic activity of the cortex. *J. cell. comp. Physiol.*, 1936, 8, 41–62.
3. ——. Subjective brightness in relation to flash rate and the light-dark ratio. *J. exp. Psychol.*, 1938, 23, 313–319.
4. ——. Some factors in brightness discrimination. Psychol. Rev., 1939, 46, 337–358.
5. ——. Some effects of intermittent photic stimulation. *J. exp. Psychol.*, 1939, 25, 462–480.
6. ——. The relation between cortical response to visual stimulation and changes in the alpha rhythm. *J. exp. Psychol.*, 1940, 27, 624–639.
7. ——. *Vision, a study of its basis.* New York: Van Nostrand Co., 1941. 350 pp.
8. ——. Visual sensation and its dependence on the neurophysiology of the optic pathway. *Biol. Sympos.*, 1942, 7, 87–106.
9. Bishop, G. H. Cyclic changes in the excitability of the optic pathway. *Amer. J. Physiol.*, 1933, 103, 213–224.
10. Briggs, G. E. Photosynthesis in intermittent light in relation to current formulations of the principles of the photosynthetic mechanism. *Biol. Rev.*, 1935, 10, 460–482.
11. Brown, H. T., & Escombe, F. Researches on some of the physiological processes of green leaves, with special reference to the interchange of energy between the leaf and its surroundings. *Proc. roy. Soc.*, London. Ser. B., 1905, 76, 29–111.
12. Brücke, E. Über den Nutzeffect intermittirender Netzhautreizungen. *S. B. Akad. Wiss. Wien (math. naturw. Classe)*, 1864, 49, 128–153.
13. Burke, D., & Lineweaver, H. The kinetic mechanism of photosynthesis. *Cold Spring Harb. Sympos. quant. Biol.*, 1935, 3, 165–183.
14. Davis, R. Experimental study of the relation between intermittent and non-intermittent sector-wheel photographic exposures. *Sci. Papers, U. S. Bureau of Standards*, 1926, 21, 95–139.

15. Dolley, W. L., Jr. The factors involved in stimulation by intermittent light. *J. exp. Zool.*, 1932, 62, 319–326.
16. Emerson, R., & Arnold, W. A separation of the reactions in photosynthesis by means of intermittent light. *J. gen. Physiol.*, 1932, 15, 391–420.
17. Hecht, S. Intensity discrimination and the stationary state. *J. gen. Physiol.*, 1924, 6, 355–373.
18. ——. Rods, cones, and the chemical basis of vision. *Physiol. Rev.*, 1937, 17, 239.
19. ——, Schlaer, S., & Smith, E. Nature of intermittent stimulation by light. *Cold Spring Harb. Sympos. quant. Biol.*, 1935, 3, 237–244.
20. ——, & Verrijp, C. D. Intermittent stimulation by light. IV. A theoretical interpretation of the quantitative data of flicker, *J. gen. Physiol.*, 1933, 17, 269–282.
21. ——, & Wolf, E. Intermittent stimulation by light. I. The validity of Talbot's law for Mya. *J. gen. Physiol.*, 1932, 15, 369–389.
22. Jahn, T. L., & Wulff, V. J. Effect of temperature upon the retinal action potential. *J. cell. comp. Physiol.*, 1943, 21, 41–51.
23. Mast, S. O. Photic orientation in insects with special reference to the drone-fly, Eristalis tenax and the robber-fly, Erax rufibarbis. *J. exp. Zool.*, 1923, 38, 109–205.
24. Warburg, O. Über die Geschwindigkeit der photochemischen Kohlensaurezersetzung in lebenden Zellen. Parts I and II, *Biochem. Z.*, 1919, 100, 230–270; 1920, 103, 188–217.
25. Wulff, V. J. Correlation of photochemical events with the action potential of the retina. *J. cell. comp. Physiol.*, 1943, 21, 319–326.

AMERICAN PSYCHOLOGICAL PERIODICALS

American Journal of Psychology—Ithaca, N. Y.; Cornell University. Subscription $6.50. 624 pages annually. Edited by K. M. Dallenbach, Madison Bentley, and E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

Journal of Genetic Psychology—Provincetown, Mass.; The Journal Press. Subscription $14.00 per annum (2 volumes). 1000 pages annually. Edited by Carl Murchison. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

Psychological Review—Northwestern University, Evanston, Illinois; American Psychological Association, Inc. Subscription $5.50. 540 pages annually. Edited by Herbert S. Langfeld. Bi-monthly. General psychology. Founded 1894.

Psychological Monographs—Northwestern University, Evanston, Illinois; American Psychological Association, Inc. Subscription $6.00 per volume. 500 pages. Edited by John F. Dashiell. Without fixed dates, each issue one or more researches. Founded 1895.

Psychological Bulletin—Northwestern University, Evanston, Illinois; American Psychological Association, Inc. Subscription $7.00. 665 pages annually. Edited by John E. Anderson. Monthly (10 numbers). Psychological literature. Founded 1904.

Archives of Psychology—New York, N. Y.; Columbia University. Subscription $6.00 per volume. 500 pages. Edited by R. S. Woodworth. Without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Northwestern University, Evanston, Illinois; American Psychological Association, Inc. Subscription $5.00. 560 pages annually. Edited by Gordon W. Allport. Quarterly. Founded 1906.

Journal of Educational Psychology—Baltimore, Md.; Warwick & York. Subscription $6.00. 720 pages annually. Edited by J. W. Dunlap, P. M. Symonds, and H. E. Jones. Monthly except June to August. Founded 1910.

Psychoanalytic Review—New York, N. Y.; 64 West 56th St. Subscription $6.00. 500 pages annually. Edited by Smith Ely Jelliffe. Quarterly. Founded 1913.

Journal of Experimental Psychology—Northwestern University, Evanston, Illinois; American Psychological Association, Inc. Subscription $7.00 per annum. 520 pages annually. Edited by Samuel W. Fernberger. Bi-Monthly. Founded 1916.

Journal of Applied Psychology—Northwestern University, Evanston, Illinois; American Psychological Association, Inc. Subscription $6.00. 480 pages annually. Edited by Donald G. Paterson. Bi-monthly. Founded 1917.

Journal of Comparative Psychology—Baltimore, Md.; Williams & Wilkins Co. Subscription $14.00 per annum (2 volumes). 1000 pages annually. Edited by Roy M. Dorcus. Bi-monthly. Founded 1921.

Comparative Psychology Monographs—Baltimore, Md.; Williams & Wilkins Co. Subscription $6.00 per volume. 400 pages. Edited by Roy M. Dorcus. Without fixed dates, each number a single research. Founded 1922.

Genetic Psychology Monographs—Provincetown, Mass.; The Journal Press. Subscription $7.00. 500 pages annually. Edited by Carl Murchison. Bi-monthly. Each number one complete research. Child behavior, animal behavior, and comparative psychology. Founded 1925.

Psychological Abstracts—Northwestern University, Evanston, Illinois; American Psychological Association, Inc. Subscription $7.00. 700 pages annually. Edited by Walter S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

Journal of General Psychology—Provincetown, Mass.; The Journal Press. Subscription $14.00 per annum (2 volumes). 1000 pages annually. Edited by Carl Murchison. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

Journal of Social Psychology—Provincetown, Mass.; The Journal Press. Subscription $7.00. 500 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.

Psychoanalytic Quarterly—Albany, N. Y.; 372–374 Broadway. Subscription $6.00. 560 pages annually. Edited by Bertram D. Lewin and others. Quarterly. Founded 1932.

Character and Personality—Durham, N. C.; Duke University Press. Subscription $2.00. 360 pages annually. Edited by Karl Zener. Quarterly. Founded 1932.

Journal of Psychology—Provincetown, Mass.; The Journal Press. Subscription $14.00 per annum (2 volumes). 800–1200 pages annually. Edited by Carl Murchison. Quarterly. Founded 1936.

Psychometrika—University of Chicago, Chicago, Ill.; Psychometric Society. Subscription $10.00. 320 pages annually. Edited by L. L. Thurstone and others. Quarterly. Quantitative methods in psychology. Founded 1936.

Psychological Record—Bloomington, Ind.; Principia Press. Subscription $4.00. 500 pages annually. Edited by J. R. Kantor and C. M. Louttit. Without fixed dates, each number a single research. General psychology. Founded 1937.

Journal of Consulting Psychology—Lancaster, Penn.; Science Printing Co. Subscription $3.00. 240 pages annually. Edited by J. P. Symonds. Bi-monthly. Founded 1937.